(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 25 May 2001 (25.05.2001)

PCT

(10) International Publication Number WO 01/35725 A1

- (51) International Patent Classification⁷: A01H 1/00, 5/00, C12N 5/04, 15/00, 15/82, C12P 21/02
- (21) International Application Number: PCT/US00/31414
- (22) International Filing Date:

14 November 2000 (14.11.2000)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

 60/166,228
 17 November 1999 (17.11.1999)
 US

 60/197,899
 17 April 2000 (17.04.2000)
 US

 60/227,439
 22 August 2000 (22.08.2000)
 US

- (71) Applicant (for all designated States except US): MENDEL BIOTECHNOLOGY, INC. [US/US]; 21375 Cabot Boulevard, Hayward, CΛ 94541 (US).
- (71) Applicants and
- (72) Inventors: JIANG, Cai-Zhong [CN/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US). HEARD, Jacqueline [US/US]; 810 Guildford Avenue, San Mateo, CA 94402 (US). PINEDA, Omaira [US/US]; 19563 Helen Place, Castro Valley, CA 94546 (US). PILGRIM, Marsha [—/US]; 790 Saltillo Place, Fremont, CA 94536 (US). ADAM, Luc [CA/US]; 25800 Industrial Boulevard, L403, Hayward, CA 94545 (US). RIECHMANN, Jose, Luis [ES/US]; 115 Moss Avenue #308, Oakland,

CA 94611 (US). YU, Guo-Liang [CN/US]; 242 Gravatt Drive, Berkeley, CA 94705 (US). SAMAHA, Raymond [US/US]; 2224 Albert Lane, Capitola, CA 95010 (US).

- (74) Agent: GUERRERO, Karen; Mendel Biotechnology, Inc., 21375 Cabot Boulevard, Hayward, CA 94545 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- With international search report.
- Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



1/35725 A

(54) Title: YIELD-RELATED GENES

(57) Abstract: Recombinant polynucleotides and methods for modifying the phenotype of a plant are provided. In particular, the phenotype that is being modified is a plant's sugar-sensing characteristics.

YIELD-RELATED GENES

RELATED APPLICATION INFORMATION

The present invention claims the benefit from US Provisional Patent Application Serial Nos. 60/166,228 filed November 17, 1999 and 60/197,899 filed April 17, 2000 and "Plant Trait Modification III" filed August 22, 2000.

FIELD OF THE INVENTION

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for phenotypically modifying a plant.

10 BACKGROUND OF THE INVENTION

5

15

20

25

30

Because sugars are important signaling molecules, the ability to control either the concentration of a signaling sugar or how the plant perceives or responds to a signaling sugar can be used to control plant development, physiology or metabolism. For example, the flux of sucrose (a disaccharide sugar used for systemically transporting carbon and energy in most plants) has been shown to affect gene expression and alter storage compound accumulation in seeds (Wobus et al (1999) Biol. Chem. 380:937-944). Manipulation of the sucrose signaling pathway in seeds may therefore cause seeds to have more protein, oil or carbohydrate, depending on the type of manipulation. Similarly, in tubers, sucrose is converted to starch which is used as an energy store.

It is thought that sugar signaling pathways may partially determine the levels of starch synthesized in the tubers (Zrenner et al. (1996) Plant J. 9:671-681). The manipulation of sugar signaling in tubers could lead to tubers with a higher starch content. Thus, manipulating the sugar signal transduction pathway may lead to altered gene expression to produce plants with desirable traits. In particular, manipulation of sugar signal transduction pathways could be used to alter source-sink relationships in seeds, tubers, roots and other storage organs leading to an increase in yield.

The present invention provides novel transcription factors useful for modifying a plant's phenotype in desirable ways by modifying a plant's sugar-sensing characteristics and thereby, increasing the yield.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID Nos. 2N, where N=1-35, or a complementary nucleotide sequence thereof; (b) a nucleotide sequence encoding a polypeptide

comprising a conservatively substituted variant of a polypeptide of (a); (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID Nos. 2N-1, where N=1-35, or a complementary nucleotide sequence thereof; (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c); (e) a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence of one or more of: (a), (b), (c), or (d); (f) a nucleotide sequence comprising at least 15 consecutive nucleotides of a sequence of any of (a)-(e); (g) a nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide having a biological activity that modifies a plant's sugar-sensing characteristics; (h) a nucleotide sequence having at least 34% sequence identity to a nucleotide sequence of any of (a)-(g); (i) a nucleotide sequence having at least 60% identity sequence identity to a nucleotide sequence of any of (a)-(g); (j) a nucleotide sequence which encodes a polypeptide having at least 34% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-35; (k) a nucleotide sequence which encodes a polypeptide having at least 60% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-35; and (1) a nucleotide sequence which encodes a conserved domain of a polypeptide having at least 65% sequence identity to a conserved domain of a polypeptide of SEQ ID Nos. 2N, where N=1-35. The recombinant polynucleotide may further comprise a constitutive, inducible, or tissue-active promoter operably linked to the nucleotide sequence. The invention also relates to compositions comprising at least two of the above described polynucleotides.

10

15

20

25

30

In a second aspect, the invention is an isolated or recombinant polypeptide comprising a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotide described above.

In another aspect, the invention is a transgenic plant comprising one or more of the above described recombinant polynucleotides. In yet another aspect, the invention is a plant with altered expression levels of a polynucleotide described above or a plant with altered expression or activity levels of an above described polypeptide. Further, the invention is a plant lacking a nucleotide sequence encoding a polypeptide described above. The plant may be a soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, or vegetable brassicas plant.

In a further aspect, the invention relates to a cloning or expression vector comprising the isolated or recombinant polynucleotide described above or cells comprising the cloning or expression vector.

In yet a further aspect, the invention relates to a composition produced by incubating a polynucleotide of the invention with a nuclease, a restriction enzyme, a polymerase; a polymerase and a primer; a cloning vector, or with a cell.

5

10

15

20

25

30

Furthermore, the invention relates to a method for producing a plant having improved sugar-sensing traits. The method comprises altering the expression of an isolated or recombinant polynucleotide of the invention or altering the expression or activity of a polypeptide of the invention in a plant to produce a modified plant, and selecting the modified plant for modified sugar-sensing traits.

In another aspect, the invention relates to a method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of the invention. The method comprises expressing a polypeptide encoded by the polynucleotide in a plant; and identifying at least one factor that is modulated by or interacts with the polypeptide. In one embodiment the method for identifying modulating or interacting factors is by detecting binding by the polypeptide to a promoter sequence, or by detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system, or by detecting expression of a factor by hybridization to a microarray, subtractive hybridization or differential display.

In yet another aspect, the invention is a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest. The method comprises placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of the invention and monitoring one or more of the expression level of the polynucleotide in the plant, the expression level of the polypeptide in the plant, and modulation of an activity of the polypeptide in the plant.

In yet another aspect, the invention relates to an integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of the invention, or to a polypeptide encoded by the polynucleotide. The integrated system, computer or computer readable medium may comprise a link between one or more sequence strings to a modified plant sugar-sensing trait.

In yet another aspect, the invention is a method for identifying a sequence similar or homologous to one or more polynucleotides of the invention, or one or more polypeptides encoded by the polynucleotides. The method comprises providing a sequence database; and, querying the sequence database with one or more target sequences corresponding to the one or

more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

The method may further comprise of linking the one or more of the polynucleotides of the invention, or encoded polypeptides, to a modified plant sugar-sensing phenotype.

5

10

15

20

25

30

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a table of exemplary polynucleotide and polypeptide sequences of the invention. The table includes from left to right for each sequence: the SEQ ID No., the internal code reference number (GID), whether the sequence is a polynucleotide or polypeptide sequence, and identification of any conserved domains for the polypeptide sequences.

Figure 2 provides a table of exemplary sequences that are homologous to other sequences provided in the Sequence Listing and that are derived from *Arabidopsis thaliana*. The table includes from left to right: the SEQ ID No., the internal code reference number (GID), identification of the homologous sequence, whether the sequence is a polynucleotide or polypeptide sequence, and identification of any conserved domains for the polypeptide sequences.

Figure 3 provides a table of exemplary sequences that are homologous to the sequences provided in Figures 1 and 2 and that are derived from plants other than *Arabidopsis thaliana*. The table includes from left to right: the SEQ ID No., the internal code reference number (GID), the unique GenBank sequence ID No. (NID), the probability that the comparison was generated by chance (P-value), and the species from which the homologous gene was identified.

DETAILED DESCRIPTION

The present invention relates to polynucleotides and polypeptides, e.g. for modifying phenotypes of plants.

In particular, the polynucleotides or polypeptides are useful for modifying traits associated with a plant's sugar-sensing characteristics when the expression levels of the polynucleotides or expression levels or activity levels of the polypeptides are altered. Sugars are central regulatory molecules that control aspects of physiology, metabolism and development. Therefore, the polynucleotides and polypeptides are useful for modifying the growth and germination rates of plants, photosynthesis, glyoxylate metabolism, respiration, starch and

sucrose synthesis and degradation, pathogen response, wounding response, cell cycle regulation, pigmentation, flowering and senescense of plants and for modifying sink-source relationships in seeds, tubers, roots and other storage organs leading to an increase in yield.

5

10

15

20

25

30

The polynucleotides of the invention encode plant transcription factors. The plant transcription factors are derived, e.g., from Arabidopsis thaliana and can belong, e.g., to one or more of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) J. Biol. Chem. 379:633-646); the MYB transcription factor family (Martin and Paz-Ares (1997) Trends Genet. 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) J. Biol. Chem. 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) Mol. Gen. Genet. 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) Plant Cell 4:1575-1588); the miscellaneous protein (MISC) family (Kim et al. (1997) Plant J. 11:1237-1251); the zinc finger protein (Z) family (Klug and Schwabe (1995) FASEB J. 9: 597-604); the homeobox (HB) protein family (Duboule (1994) Guidebook to the Homeobox Genes, Oxford University Press); the CAAT-element binding proteins (Forsburg and Guarente (1989) Genes Dev. 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) Mol. Gen. Genet. 1996 250:7-16); the NAM protein family; the IAA/AUX proteins (Rouse et al. (1998) Science 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) Prot. Profile 1:639-709); the DNAbinding protein (DBP) family (Tucker et al. (1994) EMBO J. 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) FASEB J. 8:192-200); the BPF-1 protein (Box Pbinding factor) family (da Costa e Silva et al. (1993) Plant J. 4:125-135); and the golden protein (GLD) family (Hall et al. (1998) Plant Cell 10:925-936).

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e, expression) of proteins; as regulators of plant gene expression, as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like, of as substrates for cloning e.g., including digestion or ligation reactions, and for identifying exogenous or endogenous modulators of the transcription factors.

DEFINITIONS

10

15

20

25

30

A "polynucleotide" is a nucleic acid sequence comprising a plurality of polymerized nucleotide residues, e.g., at least about 15 consecutive polymerized nucleotide residues, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. In many instances, a polynucleotide comprises a nucleotide sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

The term "transgenic plant" refers to a plant that contains genetic material, not found in a wild type plant of the same species, variety or cultivar. The genetic material may include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation.

5

10

15

20

25

30

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after transformation of a parent plant. A plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

The phrase "ectopically expression or altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild type plant or a reference plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern can be transient or stable, constitutive or inducible. In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

The term "fragment" or "domain," with respect to a polypeptide, refers to a subsequence of the polypeptide. In some cases, the fragment or domain, is a subsequence of the polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that binds to a DNA promoter region, an activation domain or a domain

for protein-protein interactions. Fragments can vary in size from as few as 6 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. In reference to a nucleotide sequence, "a fragment" refers to any subsequence of a polynucleotide, typically, of at least consecutive about 15 nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50, of any of the sequences provided herein.

5

10

15

20

25

30

The term "trait" refers to a physiological, morphological, biochemical or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by available biochemical techniques, such as the protein, starch or oil content of seed or leaves or by the observation of the expression level of genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield or pathogen tolerance.

"Trait modification" refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference), at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild type plant.

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenyllipids (such as chlorophylls and carotenoids),

glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

POLYPEPTIDES AND POLYNUCLEOTIDES OF THE INVENTION

5

10

15

20

25

30

The present invention provides, among other things, transcription factors (TFs), and transcription factor homologue polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides. These polypeptides and polynucleotides may be employed to modify a plant's sugar-sensing characteristics.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening Arabidopsis thaliana and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5' and 3' ends. The full length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5' and 3' ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention were ectopically expressed in overexpressor or knockout plants and changes in the sugar-sensing characteristics of the plants were observed.

Therefore, the polynucleotides and polypeptides can be employed to improve the sugar-sensing characteristics of plants.

Making polynucleotides

5

10

15

20

25

30

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homologue polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homologue polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homologue polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, <u>Guide to Molecular Cloning Techniques</u>, <u>Methods in Enzymology</u> volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al., <u>Molecular Cloning - A Laboratory Manual</u> (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and <u>Current Protocols in Molecular Biology</u>, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger, Sambrook, and Ausubel, as well as Mullis et al., (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are

summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, e.g., Ausubel, Sambrook and Berger, all supra.

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100 bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucletotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) Tetrahedron Letters 22:1859-69; and Matthes et al. (1984) EMBO J. 3:801-5. According to such methods, oligonucleotides are synthesized, purified, annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

HOMOLOGOUS SEQUENCES

5

10

15

20

25

30

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from Arabidopsis thaliana or from other plants of choice are also an aspect of the invention. Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, brussel sprouts and kohlrabi). Other crops, fruits and vegetables whose phenotype can be changed include barley, rye, millet, sorghum, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato, and beans. The homologous sequences may also be derived from woody species, such pine, poplar and eucalyptus.

Transcription factors that are homologous to the listed sequences will typically share at least about 34% amino acid sequence identity. More closely related transcription factors can share at least about 50%, about 60%, about 65%, about 70%, about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity.

Identifying Nucleic Acids by Hybridization

10

15

20

25

30

Polynucleotides homologous to the sequences illustrated in the Sequence Listing can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number), as described in more detail in the references cited above.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is about 5°C to 20°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions, e.g., to a unique subsequence, of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C, for example 0.2 x SSC, 0.1% SDS at 65° C. For identification of less closely related homologues washes can

be performed at a lower temperature, e.g., 50° C. In general, stringency is increased by raising the wash temperature and/or decreasing the concentration of SSC.

As another example, stringent conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing date of the application. Conditions can be selected such that a higher signal to noise ratio is observed in the particular assay which is used, e.g., about 15x, 25x, 35x, 50x or more. Accordingly, the subject nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. Again, higher signal to noise ratios can be selected, e.g., about 5x, 10x, 25x, 35x, 50x or more. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radio active label, or the like.

Alternatively, transcription factor homologue polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homologue nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homologue, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologues, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

SEQUENCE VARIATIONS

5

10

15

20

25

30

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homologue polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing.

For example, Table 1 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 1

5

10

15

20

Amino acids			Codon					
Alanine	Ala	Α	GCA	GCC	GCG	GCU		
Cysteine	Cys	С	TGC	TGT				
Aspartic acid	Asp	D	GAC	GAT				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	TTC	TTT				
Glycine	Gly	G	GGA	GGC	GGG	GGT		
Histidine	His	H	CAC	CAT				
Isoleucine	Ile	I	ATA	ATC	ATT			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT
Methionine	Met	M	ATG					
Asparagine	Asn	N	AAC	AAT				
Proline	Pro	P	CCA	CCC	CCG	CCT		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT
Threonine	Thr	T	ACA	ACC	ACG	ACT		
Valine	Val	V	GTA	GTC	GTG	GTT		
Tryptophan	Trp	W	TGG					
Tyrosine	Tyr	Y	TAC	TAT				

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) Meth. Enzymol. (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid

substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 2

Residue	Conservative Substitutions				
Ala	Ser				
Arg	Lys				
Asn	Gln; His				
Asp	Glu				
Gln	Asn				
Cys	Ser				
Glu	Asp				
Gly	Pro				
His	Asn; Gln				
Ile	Leu, Val				
Leu	Ile; Val				
Lys	Arg; Gln				
Met	Leu; Ile				
Phe	Met; Leu; Tyr				
Ser	Thr; Gly				
Thr	Ser;Val				
Trp	Тут				
Tyr	Trp; Phe				
Val	Ile; Leu				

Substitutions that are less conservative than those in Table 2 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

5

10

FURTHER MODIFYING SEQUENCES OF THE INVENTION—MUTATION/ FORCED EVOLUTION

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

5

10

15

20

25

30

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, e.g., by Stemmer (1994) Nature 370:389-391, and Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for S. cerevisiae and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and E. coli prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51; 113-119) and synthetic peptides (Giniger and Ptashne, (1987) Nature 330:670-672).

EXPRESSION AND MODIFICATION OF POLYPEPTIDES

5

10

15

20

25

30

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homologue.

Vectors, Promoters and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts which describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant topics, include Berger, Sambrook and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology, Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of Agrobacterium tumefaciens, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technology 3: 637-642, for dicotyledonous plants.

10

15

20

25

30

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9: 957-962) and com (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48, and for Agrobacterium-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (see, e.g., Odel et al. (1985) Nature 313:810); the nopaline synthase promoter (An et al. (1988) Plant Physiol 88:547); and the octopine synthase promoter (Fromm et al. (1989) Plant Cell 1: 977).

5

10

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorable be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the dru 1 promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) Plant Mol Biol 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) Plant Mol Biol 37:977-988), flower-15 specific (Kaiser et al, (1995) Plant Mol Biol 28:231-243), pollen (Baerson et al. (1994) Plant Mol Biol 26:1947-1959), carpels (Ohl et al. (1990) Plant Cell 2:837-848), pollen and ovules (Baerson et al. (1993) Plant Mol Biol 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) Plant Mol Biol 39:979-990 or Baumann et al. (1999) Plant Cell 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) Plant Mol Biol 38:743-753), 20 promoters responsive to gibberellin (Shi et al. (1998) Plant Mol Biol 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) Plant Mol Biol 22: 13-23), light (e.g., the pea rbcS-3A promoter, Kuhlemeier et al. (1989) Plant Cell 1:471, and the maize rbcS promoter, Schaffner and Sheen (1991) Plant Cell 3: 997); wounding (e.g., wunI, Siebertz et al. (1989) Plant Cell 1: 961); 25 pathogens (such as the PR-1 promoter described in Buchel et al. (1999) Plant Mol. Biol. 40:387-396, and the PDF1.2 promoter described in Manners et al. (1998) Plant Mol. Biol. 38:1071-80), and chemicals such as methyl jasmonate or salicylic acid (Gatz et al. (1997) Plant Mol Biol 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) Science 270: 1986-1988); or late seed development 30 (Odell et al. (1994) Plant Physiol 106:447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant

genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

5

10

15

20

25

30

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e, nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acids, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook and Ausubel.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al., (1985) Proc. Natl. Acad. Sci. USA 82, 5824, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., (1982) Molecular Biology of Plant Tumors, (Academic Press, New York) pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327, 70-73), use of pollen as vector (WO 85/01856), or use of Agrobacterium

tumefaciens or A. rhizogenes carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by Agrobacterium tumefaciens, and a portion is stably integrated into the plant genome (Horsch et al. (1984) Science 233:496-498; Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80, 4803).

5

The cell can include a nucleic acid of the invention which encodes a polypeptide, wherein the cells expresses a polypeptide of the invention. The cell can also include vector sequences, or the like. Furthermore, cells and transgenic plants which include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

10

15

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

Modified Amino Acids

20

Polypeptides of the invention may contain one or more modified amino acids. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

25

30

Non-limiting examples of a modified amino acid include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids, biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acids are replete throughout the literature.

IDENTIFICATION OF ADDITIONAL FACTORS

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phentoype or trait of

interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream gene 5 with which is subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homologue of the invention is expressed in a host cell, e.g, a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional 10 gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (i.e., binding sites) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific 15 nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. 20 (1999) Nature Biotechnology 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or-heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

25

30

The two-hybrid system detects protein interactions in vivo and is described in Chien, et al., (1991), Proc. Natl. Acad. Sci. USA 88, 9578-9582 and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid

and the cDNA library are transformed into a strain of the yeast Saccharomyces cerevisiae that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

10 IDENTIFICATION OF MODULATORS

5

15

20

25

30

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northerns, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) <u>Current Protocols in Molecular Biology</u>, John Wiley & Sons (1998). Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity. Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

5

10

15

20

25

30

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) Nature Biotechnology, 14(3):309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. Science (1996) 274:1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175, Furka, Int. J. Pept. Prot. Res. 37:487-493 (1991) and Houghton et al. Nature 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high throughput screening systems are commercially available.

These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s)

appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be incubated with cells/plants/ etc. in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA/ protein expression, etc., according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators which inhibit transcriptional activation or transcriptional repression. Either expression of the nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention.

SUBSEQUENCES

5

10

15

20

25

30

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least

20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra- high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted supra.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologues of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook and Ausubel, supra.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

PRODUCTION OF TRANSGENIC PLANTS

5

10

15

20

25

30

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors (or transcription factor homologues) of the invention, as compared with the levels of the same protein found in a wild type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved sugar-sensing characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Antisense and Cosuppression Approaches

5

10

15

20

25

30

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997)

Antisense Technology: A Practical Approach IRL Press at Oxford University, Oxford, England. In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a "knock-out") of a transcription factor or transcription factor homologue polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homologue cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous

mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Vectors in which RNA encoded by a transcription factor or transcription factor homologue cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No. 5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

10

15

20

25

30

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating it's activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) Genes and Development 13: 139-141).

Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of Agrobacterium tumefaciens. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homologue gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation (Koncz et al. (1992) Methods in Arabidopsis Research, World Scientific).

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homologue, e.g., by homologous recombination (Kempin et al. (1997) Nature 389:802).

A plant trait can also be modified by using the cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites is excised. If the lox sites are in the opposite orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of

the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) Nature 390 698-701; Kakimoto et al. (1996) Science 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (See, e.g., PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA binding specificity of zinc finger proteins by changing particular amino acids in the DNA binding motif).

5

10

15

20

25

30

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homologue, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledenous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip), *Cruciferae* (cabbage, radish, rapeseed, broccoli, etc.), *Curcurbitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) <u>Handbook of Plant Cell Culture – Crop Species.</u> Macmillan Publ. Co. Shimamoto et al. (1989) <u>Nature</u> 338:274-276; Fromm et al. (1990) <u>Bio/Technology</u> 8:833-839; and Vasil et al. (1990) <u>Bio/Technology</u> 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated

transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and Agrobacterium tumeficiens mediated transformation. Transformation means introducing a nucleotide sequence in a plant in a manner to cause stable or transient expression of the sequence.

5

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

10

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

15

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

20

INTEGRATED SYSTEMS—SEQUENCE IDENTITY

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved sugar-sensing characteristics, with one or more identified sequence.

30

25

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madision, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PhytoSeq (Incyte Pharmaceuticals, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology alignrithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85: 2444, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al., supra.

5

10

15

20

25

30

A variety of methods of determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. J. Mol. Biol 215:403-410 (1990). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an

expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

5

10

15

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see*, e.g., Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element which displays an alignment produced by the alignment instruction set.

25

20

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, e.g. through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

30

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is

provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

10 EXAMPLES

5

25

30

The following examples are intended to illustrate but not limit the present invention.

EXAMPLE I. FULL LENGTH GENE IDENTIFICATION AND CLONING

Putative transcription factor sequences (genomic or ESTs) related to known

transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, Arabidopsis *thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ³²P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO₄ pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60 °C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSCC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the MarathonTM cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded

cDNA, blunting cDNA ends, followed by ligation of the MarathonTM Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

EXAMPLE II. CONSTRUCTION OF EXPRESSION VECTORS

5

10

15

20

25

30

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20 or pMEN65, which are both derived from pMON316 (Sanders et al, (1987) Nucleic Acids Research 15:1543-58) and contain the CaMV 35S promoter to express transgenes. To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with SalI and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, CA). The fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma).

Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen, CA).

EXAMPLE III. TRANSFORMATION OF AGROBACTERIUM WITH THE EXPRESSION VECTOR

After the plasmid vector containing the gene was constructed, the vector was used to transform Agrobacterium tumefaciens cells expressing the gene products. The stock of Agrobacterium tumefaciens cells for transformation were made as described by Nagel et al. (1990) FEMS Microbiol Letts. 67: 325-328. Agrobacterium strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A₆₀₀) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4°C. Cells were then

resuspended in 250 μ l chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 μ l chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 μ l and 750 μ l, respectively. Resuspended cells were then distributed into 40 μ l aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

5

10

15

30

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 μl of Agrobacterium cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 μF and 200 μF using a Gene Pulser II apparatus (Bio-Rad). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28°C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 μg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28°C. Single colonies were then picked and inoculated in fresh medium. The presence of the plasmid construct was verified by PCR amplification and sequence analysis.

EXAMPLE IV. TRANSFORMATION OF ARABIDOPSIS PLANTS WITH AGROBACTERIUM TUMEFACIENS WITH EXPRESSION VECTOR

After transformation of Agrobacterium tumefaciens with plasmid vectors

containing the gene, single Agrobacterium colonies were identified, propagated, and used to transform Arabidopsis plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were iñoculated with the colonies and grown at 28° C with shaking for 2 days until an absorbance (A₆₀₀) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1

X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μM benzylamino purine (Sigma), 200 μl/L Silwet L-77 (Lehle Seeds) until an absorbance (A₆₀₀) of 0.8 was reached.

Prior to transformation, Arabidopsis thaliana seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 μ E/m²/sec) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of Agrobacterium infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

EXAMPLE V. IDENTIFICATION OF ARABIDOPSIS PRIMARY TRANSFORMANTS

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H₂O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H₂O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 μE/m²/sec) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T₁ generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T₂) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants varies from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

30

5

10

15

20

EXAMPLE VI. IDENTIFICATION OF ARABIDOPSIS PLANTS WITH TRANSCRIPTION FACTOR GENE KNOCKOUTS

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al (1999) Plant Cell 11:2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

15 EXAMPLE VII. IDENTIFICATION OF SUGAR-SENSING CHARACTERISTICS PHENOTYPE IN OVEREXPRESSOR OR GENE KNOCKOUT PLANTS

Experiments were performed to identify those transformants or knockouts that exhibited modified sugar-sensing. For such studies, seeds from transformants were germinated on media containing 5% glucose or 9.4% sucrose which normally partially restrict hypocotyl elongation. Plants with altered sugar sensing may have either longer or shorter hypocotyls than normal plants when grown on this media. Additionally, other plant traits may be varied such as root mass.

Table 3 shows the phenotypes observed for particular overexpressor or knockout plants and provides the SEQ ID No., the internal reference code (GID), whether a knockout or overexpressor plant was analyzed and the observed phenotype.

25

20

5

10

Table 3

SEQ ID No.	GID	Knockout (OE) or overexpressor KO)	Phenotype observed
1	G26	OE	Decreased germination and growth on glucose medium
3	G38	OE	Reduced germination on glucose medium
5	G43	OE	Decreased germination and growth on glucose medium
7	G207	OE	Decreased germination on glucose medium
9	G241	OE	Decreased germination and growth on glucose medium
11	G254	OE	Decreased germination and growth on glucose medium
13	G263	OE	Decreased root growth on sucrose medium
15	G308	OE	No germination on glucose medium
17	G536	OE	Decreased germination and growth on glucose medium
19	G680	OE	Reduced germination on glucose medium
21	G867	OE	Better seedling vigor on sucrose medium
23	G912	OE	Reduced cotyledon expansion in glucose
25	G996	OE	Reduced germination on glucose medium
27	G1068	OE	Reduced cotyledon expansion in glucose
29	G1337	OE	Decreased germination on sucrose medium

For a particular overexpressor that shows a less beneficial sugar-sensing characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial sugar-sensing characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

EXAMPLE VIII. IDENTIFICATION OF HOMOLOGOUS SEQUENCES

5

10

15

20

Homologous sequences from *Arabidopsis* and plant species other than *Arabidopsis* were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) <u>J. Mol. Biol.</u> 215:403-410; and Altschul et al. (1997) <u>Nucl. Acid Res.</u> 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) <u>Proc. Natl. Acad. Sci. USA</u> 89: 10915-10919).

Identified Arabidopsis homologous sequences are provided in Figure 2 and included in the Sequence Listing. The percent sequence identity among these sequences is as low as 47% sequence identity. Additionally, the entire NCBI GenBank database was filtered for sequences from all plants except Arabidopsis thaliana by selecting all entries in the NCBI GenBank database associated with NCBI taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (Arabidopsis thaliana). These sequences were compared to sequences representing genes of SEQ IDs Nos. 1-54 on 9/26/2000 using the Washington University TBLASTX algorithm (version 2.0a19MP). For each gene of SEQ IDs

Nos. 1-54, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of 3.6e-40 is 3.6×10^{-40} . For up to ten species, the gene with the lowest P-value (and therefore the most likely homolog) is listed in Figure 3.

5

10

15

20

25

In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. The ranges of percent identity between the non-Arabidopsis genes shown in Figure 3 and the Arabidopsis genes in the sequence listing are: SEQ ID No. 1: 44%-79%; SEQ ID No. 3: 36%-72%; SEQ ID No. 5: 42%-67%; SEQ ID No. 7: 55%-82%; SEQ ID No. 9: 69%-84%; SEQ ID No. 11: 57%-90%; SEQ ID No. 13: 48%-85%; SEQ ID No. 15: 38%-85%; SEQ ID No. 17: 77%-87%; SEQ ID No. 19: 42%-88%; SEQ ID No. 21: 54%-69%; SEQ ID No. 23: 34%-71%; SEQ ID No. 25: 55%-95%; SEQ ID No. 27: 54%-95%; SEQ ID No. 29: 37%-58%; SEQ ID No. 31: 42%-70%; SEQ ID No. 33: 46%-62%; SEQ ID No. 35: 64%-84%; SEQ ID No. 37: 57%-87%; SEQ ID No. 39: 40%-80%; SEQ ID No. 41: 56%-82%; SEQ ID No. 43: 64%-93%; SEQ ID No. 45: 35%-86%; SEQ ID No. 47: 84%-91%; SEQ ID No. 49: 85%-91%; SEQ ID No. 51: 38%-89%; SEQ ID No. 53: 53%-75%; SEQ ID No. 55: 57%-72%; SEQ ID No. 57: 57%-69%; SEQ ID No. 59: 49%-86%; SEQ ID No. 61: 49%-78%; SEQ ID No. 63: 51%-86%; SEQ ID No. 65: 42%-72%; SEQ ID No. 67: 35%-69%; and SEQ ID No. 69: 36%-64%.

The polynucleotides and polypeptides in the Sequence Listing and the identified homologous sequences may be stored in a computer system and have associated or linked with the sequences a function, such as that the polynucleotides and polypeptides are useful for modifying the sugar-sensing characteristics of a plant.

All references, publications, patents and other documents herein are incorporated by reference in their entirety for all purposes. Although the invention has been described with reference to the embodiments and examples above, it should be understood that various modifications can be made without departing from the spirit of the invention.

What is claimed is:

10

15

20

1. A transgenic plant with modified sugar-sensing characteristics, which plant comprises a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID Nos. 2N, where N=1-35, or a complementary nucleotide sequence thereof;

 (b) a nucleotide sequence encoding a polypeptide comprising a conservatively substituted variant of a polypeptide of (a);
 - (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID Nos. 2N-1, where N=1-35, or a complementary nucleotide sequence thereof;
 - (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c);
 (e) a nucleotide sequence which hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a), (b), (c), or (d);
 - (f) a nucleotide sequence comprising at least 15 consecutive nucleotides of a sequence of any of (a)-(e);
 - (g) a nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide that modifies a plant's sugar-sensing characteristics;
 - (h) a nucleotide sequence having at least 34% sequence identity to a nucleotide sequence of any of (a)-(g);
 - (i) a nucleotide sequence having at least 60% identity sequence identity to a nucleotide sequence of any of (a)-(g);
 - (j) a nucleotide sequence which encodes a polypeptide having at least 34% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-35;
- 25 (k) a nucleotide sequence which encodes a polypeptide having at least 60% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-35; and (l) a nucleotide sequence which encodes a polypeptide having at least 65% sequence identity to a conserved domain of a polypeptide of SEQ ID Nos. 2N, where N=1-35.
- The transgenic plant of claim 1, further comprising a constitutive, inducible, or tissueactive promoter operably linked to said nucleotide sequence.
 - 3. The transgenic plant of claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf,

banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, and vegetable brassicas.

5

10

15

20

- 4. An isolated or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID Nos. 2N, where N=1-35, or a complementary nucleotide sequence thereof;
- (b) a nucleotide sequence encoding a polypeptide comprising a conservatively substituted variant of a polypeptide of (a);
 - (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID Nos. 2N-1, where N=1-35, or a complementary nucleotide sequence thereof;
 - (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c);
- (e) a nucleotide sequence which hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a), (b), (c), or (d);
 - (f) a nucleotide sequence comprising at least 15 consecutive nucleotides of a sequence of any of (a)-(e);
 - (g) a nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide that modifies a plant's sugar-sensing characteristics;
 - (h) a nucleotide sequence having at least 34% sequence identity to a nucleotide sequence of any of (a)-(g);
 - (i) a nucleotide sequence having at least 60% identity sequence identity to a nucleotide sequence of any of (a)-(g);
 - (j) a nucleotide sequence which encodes a polypeptide having at least 34% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-35;
 - (k) a nucleotide sequence which encodes a polypeptide having at least 60% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-35; and
- 30 (l) a nucleotide sequence which encodes a conserved domain of a polypeptide having at least 65% sequence identity to a conserved domain of a polypeptide of SEQ ID Nos. 2N, where N=1-35.

5. The isolated or recombinant polynucleotide of claim 4, further comprising a constitutive, inducible, or tissue-active promoter operably linked to the nucleotide sequence.

- 6. A cloning or expression vector comprising the isolated or recombinant polynucleotide of5 claim 4.
 - 7. A cell comprising the cloning or expression vector of claim 6.
- 8. A transgenic plant comprising the isolated or recombinant polynucleotide of claim 4.
- 9. A composition produced by one or more of:

20

25

- (a) incubating one or more polynucleotide of claim 4 with a nuclease;
- (b) incubating one or more polynucleotide of claim 4 with a restriction enzyme;
- (c) incubating one or more polynucleotide of claim 4 with a polymerase;
- (d) incubating one or more polynucleotide of claim 4 with a polymerase and a primer;
 - (e) incubating one or more polynucleotide of claim 4 with a cloning vector, or
 - (f) incubating one or more polynucleotide of claim 4 with a cell.
 - 10. A composition comprising two or more different polynucleotides of claim 4.
 - 11. An isolated or recombinant polypeptide comprising a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotide of claim 4.
 - 12. A plant ectopically expressing an isolated polypeptide of claim 11.
 - 13. A method for producing a plant having modified sugar-sensing characteristics, the method comprising altering the expression of the isolated or recombinant polynucleotide of claim 4 or the expression levels or activity of a polypeptide of claim 11 in a plant, thereby producing a modified plant, and selecting the modified plant for improved sugar-sensing characteristics thereby providing the modified plant with a modified sugar-sensing characteristics.
 - 14. The method of claim 13, wherein the polynucleotide is a polynucleotide of claim 4.

15. A method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of claim 4, the method comprising:

- (a) expressing a polypeptide encoded by the polynucleotide in a plant; and
- (b) identifying at least one factor that is modulated by or interacts with the polypeptide.

5

- 16. The method of claim 15, wherein the identifying is performed by detecting binding by the polypeptide to a promoter sequence, or detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system.
- 17. The method of claim 15, wherein the identifying is performed by detecting expression of a factor by hybridization to a microarray, subtractive hybridization or differential display.
 - 18. A method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest, the method comprising:
 - (a) placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of claim 4; and,
 - (b) monitoring one or more of:
 - (i) expression level of the polynucleotide in the plant;
 - (ii) expression level of the polypeptide in the plant;
 - (iii) modulation of an activity of the polypeptide in the plant; or
 - (iv) modulation of an activity of the polynucleotide in the plant.

20

15

- 19. An integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of claim 4, or to a polypeptide encoded by the polynucleotide.
- 20. The integrated system, computer or computer readable medium of claim 19, further comprising a link between said one or more sequence strings to a modified plant sugar-sensing characteristics phenotype.

30

- 21. A method of identifying a sequence similar or homologous to one or more polynucleotides of claim 4, or one or more polypeptides encoded by the polynucleotides, the method comprising:
 - (a) providing a sequence database; and,

(b) querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

5

- 22. The method of claim 21, wherein the querying comprises aligning one or more of the target sequences with one or more of the one or more sequence members in the sequence database.
- 10 23. The method of claim 21, wherein the querying comprises identifying one or more of the one or more sequence members of the database that meet a user-selected identity criteria with one or more of the target sequences.
- The method of claim 21, further comprising linking the one or more of the
 polynucleotides of claim 4, or encoded polypeptides, to a modified plant sugar-sensing characteristics phenotype.
 - 25. A plant comprising altered expression levels of an isolated or recombinant polynucleotide of claim 4.

20

26. A plant comprising altered expression levels or the activity of an isolated or recombinant polypeptide of claim 11.

Figure 1

SEQ ID No.	GID	cDNA or protein	conserved domain
1	G26	cDNA	
2	G26	protein	67-134
3	G38	cDNA	
4	G38	protein	76-143
5	G43	cDNA	
6	G43	protein	104-172
7	G207	cDNA	
8	G207	protein	6-106
9	G241	cDNA	
10	G241	protein	14-114
11	G254	cDNA	
12	G254	protein	62-106
13	G263	cDNA	
14	G263	protein	15-105
15	G308	cDNA	
16	G308	protein	270-274
17	G536	cDNA	
18	G536	protein	226-233
19	G680	cDNA	
20	G680	protein	24-70
21	G867	cDNA	
22	G867	protein	59-124
23	G912	cDNA	
24	G912	protein	51-118
25	G996	cDNA	
26	G996	protein	14-114
27	G1068	cDNA	
28	G1068	protein	143-150
29	G1337	cDNA	
30	G1337	protein	9-75

Figure 2

SEQ ID No.	GID	homolog	cDNA or protein	conserved domain
31	G1141	homolog of G38	cDNA	
32	G1141	homolog of G38	protein	75-142
33	G46	homolog of G43	cDNA	
34	G46	homolog of G43	protein	107-175
35	G242	homolog of G207	cDNA	
36	G242	homolog of G207	protein	6-105
37	G227	homolog of G207	cDNA	
38	G227	homolog of G207	protein	13-112
39	G1307	homolog of G241	cDNA	
40	G1307	homolog of G241	protein	14-114
41	G1327	homolog of G241	cDNA	
42	G1327	homolog of G241	protein	14-116
43	G673	homolog of G254	cDNA	
44	G673	homolog of G254	protein	37-95
45	G307	homolog of G308	cDNA	
46	G307	homolog of G308	protein	323-339
47	G529	homolog of G536	cDNA	
48	G529	homolog of G536	protein	229-236
49	G531	homolog of G536	cDNA	
50	G531	homolog of G536	protein	227-234
51	G214	homolog of G680	cDNA	
52	G214	homolog of G680	protein	22-71
53	G1930	homolog of G867	cDNA	
54	G1930	homolog of G867	protein	59-124
55	G9	homolog of G867	cDNA	
56	G9	homolog of G867	protein	62-127
57	G993	homolog of G867	cDNA	
58	G993	homolog of G867	protein	69-134
59	G41	homolog of G912	cDNA	
60	G41	homolog of G912	protein	39-106
61	G40	homolog of G912	cDNA	
62	G40	homolog of G912	protein	45-112
63	G42	homolog of G912	cDNA	
64	G42	homolog of G912	protein	48-115
65	G1127	homolog of G1068		
66	G1127	homolog of G1068		103-110, 155-162
67	G2657	homolog of G1068		
68	G2657	homolog of G1068		116-129
69	G326	homolog of G1337		
70	G326	homolog of G1337	protein	11-94, 354-400

Figure 3A

		0 1 115		0
SEQ ID No.	GID	Genbank NID	P-value	Species
1	G26	4387560		Lycopersicon esculentum
11	G26	9427282		Triticum aestivum
1	G26	7206394		Medicago truncatula
1	G26	7796858		Glycine max
1	G26	7788764		Lotus japonicus
1	G26	8098026		Hordeum vulgare
1	G26	790362		Nicotiana tabacum
1	G26	569065		Oryza sativa
1	G26	3264766		Prunus armeniaca
1	G26_	7528275		Mesembryanthemum crystallinum
3	G38	8346772		Catharanthus roseus
3	G38	7205636		Medicago truncatula
3	G38	7684799		Glycine max
3	G38_	9363798		Triticum aestivum
3	G38_	7777379_		Lotus japonicus
3	G38	8903111		Hordeum vulgare
3	G38	568076		Oryza sativa
3	G38	9434234		Lycopersicon esculentum
3	G38	7324705		Lycopersicon pennellii
3	G38	9298423		Sorghum bicolor
5	G43	5760554		Glycine max
5	G43	7778996		Lotus japonicus
5	G43	5603736	5.00E-25	Lycopersicon esculentum
5	G43	6478844		Matricaria chamomilla
5	G43	790361		Nicotiana tabacum
5	G43	7528275		Mesembryanthemum crystallinum
5	G43_	9199136		Medicago truncatula
5	G43	8980312		Catharanthus roseus
5	G43	8809570		Nicotiana sylvestris
5	G43	7627061		Gossypium arboreum
7	G207	6529807		Lycopersicon esculentum
7	G207	7564212		Medicago truncatula
7	G207	7624453		Gossypium arboreum
7	G207	5820271		Glycine max
7	G207	7322467		Lycopersicon hirsutum
7	G207	5045349		Gossypium hirsutum
7	G207	8071527		Solanum tuberosum
7	G207	7790004		Beta vulgaris
7	G207	6325768		Lotus japonicus
7	G207	286661		Oryza sativa
9	G241	6552360		Nicotiana tabacum
9	G241	6782745		Oryza sativa
9	G241	8097368		Hordeum vulgare
9	G241	20560		Petunia x hybrida
9	G241	7217727		Sorghum bicolor
9	G241	5891408		Lycopersicon esculentum
9	G241	5139803		Glycine max
9	G241_	7560175		Medicago truncatula
9	G241	8381332		Gossypium arboreum
9	G241	4886263		Antirrhinum majus
11	G254	5847380		Zea mays
11	G254_	7614730	2.90E-41	Lotus japonicus

Figure 3B

- igaio 0D				
SEQ ID No.	GID	Genbank NID	P-value	Species
11	G254	9204594		Glycine max
11	G254	9193761	6.70E-37	Medicago truncatula
11	G254	6907081		Oryza sativa
11	G254	6976741	4.30E-33	Lycopersicon esculentum
11	G254	8903196	4.20E-31	Hordeum vulgare
11	G254	9424828		Triticum aestivum
11	G254	6858452	3.40E-23	Sorghum bicolor
11	G254	3003284	0.00068	Mesembryanthemum crystallinum
13	G263	5821135		Nicotiana tabacum
13	G263	19487	7.90E-69	Lycopersicon peruvianum
13	G263	662929	5.30E-65	Glycine max
13	G263	7766273	9.20E-49	Medicago truncatula
13	G263	7720908	3.60E-42	Lotus japonicus
13	G263	9303509		Sorghum bicolor
13	G263	3326480		Gossypium hirsutum
13	G263	8107182		Lycopersicon esculentum
13	G263	8381330		Gossypium arboreum
13	G263	4528238		Citrus unshiu
15	G308	5640156		Triticum aestivum
15	G308	5640154	2.30E-134	
15	G308	6970471		Oryza sativa
15	G308	7718432		Medicago truncatula
15	G308	8330344	3.90E-76	Mesembryanthemum crystallinum
15	G308	5047560	1.50E-71	Gossypium hirsutum
15	G308	7588689		Glycine max
15	G308	7623983		Gossypium arboreum
15	G308	7780253		Lotus japonicus
15	G308	6733213		Lycopersicon esculentum
17	G536	2689478	9.50E-69	Nicotiana tabacum
17	G536	1773327	4.60E-68	Mesembryanthemum crystallinum
17	G536	2921511		Fritillaria agrestis
17	G536	1575724		Glycine max
17	G536	8515887		Populus alba x Populus tremula
17	G536	6179980		Lilium longiflorum
17	G536	1519250		Oryza sativa
17	G536	1321992		Solanum tuberosum
17	G536	7535681		Sorghum bicolor
17	G536	555973		Pisum sativum
19	G680	9258166		Glycine max
19	G680	9255178		Zea mays
19	G680	5274804		Lycopersicon esculentum
19	G680	4974199		Oryza sativa
19	G680	3325786		Gossypium hirsutum
19	G680	9119112		Medicago truncatula
19	G680	7660673		Sorghum bicolor
19	G680	7243970		Mentha x piperita
19	G680	3858093		Populus balsamifera subsp. trichocarpa
19	G680	8845091		Triticum aestivum
21	G867	7643366		Medicago truncatula
21	G867	8329389		Mesembryanthemum crystallinum
21	G867	8669779		Glycine max
21	G867	9430646		Lycopersicon esculentum
21	G867	8902194		Hordeum vulgare
1				

Figure 3C

SEQ ID No.	GID	Genbank NID	P-value	Species
21	G867	7722547		Lotus japonicus
21	G867	7324245		Lycopersicon pennellii
21	G867	8749037		Citrus x paradisi
21	G867	6069643		Oryza sativa
21	G867	9302986		Sorghum bicolor
23	G912	5616085		Brassica napus
23	G912	7410271		Lycopersicon esculentum
23	G912	7719106		Medicago truncatula
23	G912	6667103		Glycine max
23	G912	6983854		Oryza sativa
23	G912	7324530		Lycopersicon pennellii
23	G912	8904571		Triticum aestivum
23	G912	7740143		Lotus japonicus
23	G912	7644788		Pinus taeda
23	G912	5050536		Gossypium hirsutum
25	G996	7566043		Medicago truncatula
25	G996	7535969		Sorghum bicolor
25	G996	7339127		Lycopersicon esculentum
25	G996	6341619		Glycine max
25	G996	8381332		Gossypium arboreum
25	G996	5049507		Gossypium hirsutum
25	G996	6850206		Oryza sativa
		7776223		Lotus japonicus
25	G996			Hordeum vulgare
25	G996	19058		Oryza sativa subsp. indica
25	G996	4680189		Lycopersicon esculentum
27 27	G1068	7333976		Glycine max
	G1068	4405544		Zea mays
27 27	G1068	7009437 7536402		Sorghum bicolor
27	G1068			Oryza sativa
	G1068	3107210		Hordeum vulgare
27	G1068	3819186		
27	G1068	7624850		Gossypium arboreum Triticum aestivum
27	G1068	9411568		
27	G1068	5419913		Lactuca sativa
27	G1068	7721066		Lotus japonicus
29	G1337	7410432		Lycopersicon esculentum
29	G1337	3618319		Oryza sativa
29	G1337	7571599		Medicago truncatula
29	G1337	7685955		Glycine max
29	G1337	7323708		Lycopersicon hirsutum
29	G1337	4091805		Malus domestica
29	G1337	6917805		Lycopersicon pennellii
29	G1337	3341722		Raphanus sativus
29	G1337	2303680		Brassica napus
29	G1337	4557092		Pinus radiata
31	G1141	8346772		Catharanthus roseus
31	G1141	7205636		Medicago truncatula
31	G1141	7590901		Glycine max
31	G1141	7777379		Lotus japonicus
31	G1141	9363798		Triticum aestivum
31	G1141	8903111		Hordeum vulgare
31	G1141	568076		Oryza sativa
31	G1141	6527472	1.10E-17	Lycopersicon esculentum

Figure 3D

SEO ID No	CID	Conhack NID	Dueline	Cassian
SEQ ID No.	GID	Genbank NID		Species
31	G1141	7324705		Lycopersicon pennellii
31	G1141	7624302		Gossypium arboreum
33	G46	5760554		Glycine max
33	G46	7778996		Lotus japonicus
33	G46	5050094		Gossypium hirsutum
33	G46	790361		Nicotiana tabacum
33	G46	5603736		Lycopersicon esculentum
33	G46	7238955		Medicago truncatula
33	G46	8809574		Nicotiana sylvestris
33	G46	7528275		Mesembryanthemum crystallinum
33	G46	8980312		Catharanthus roseus
33	G46	6478844		Matricaria chamomilla
35	G242	6529807		Lycopersicon esculentum
35	G242	7624453		Gossypium arboreum
35	G242	7564212		Medicago truncatula
35	G242	5820271		Glycine max
35	G242	7322467		Lycopersicon hirsutum
35	G242	5045349		Gossypium hirsutum
35	G242	8071527		Solanum tuberosum
35	G242	7790004		Beta vulgaris
35	G242	7746594		Lotus japonicus
35	G242	286661		Oryza sativa
37	G227	6529807		Lycopersicon esculentum
37	G227	7624453		Gossypium arboreum
37	G227	5045349		Gossypium hirsutum
37	G227	7322467		Lycopersicon hirsutum
37	G227	5820271		Glycine max
37	G227	9199531		Medicago truncatula
37	G227	8071527		Solanum tuberosum
37	G227	7790004		Beta vulgaris
37	G227	7746594	1.70E-45	Lotus japonicus
37	G227	286661	9.20E-37	
39	G1307	8172759	2.90E-56	Medicago truncatula
39	G1307	5139807	1.60E-54	Glycine max
39	G1307	1370139	2.20E-47	Lycopersicon esculentum
39	G1307	1946264	6.20E-45	Oryza sativa
39	G1307	6552360	6.50E-45	Nicotiana tabacum
39	G1307	7500978	4.60E-39	Gossypium arboreum
39	G1307	7217727	8.70E-36	Sorghum bicolor
39	G1307	7746498	9.40E-34	Lotus japonicus
39	G1307	517491	9.90E-34	Zea mays
39	G1307	8097368		Hordeum vulgare
41	G1327	5139803		Glycine max
41	G1327	7560175		Medicago truncatula
41	G1327	6782745		Oryza sativa
41	G1327	5891408	2.30E-43	Lycopersicon esculentum
41	G1327	7217727		Sorghum bicolor
41	G1327	20560		Petunia x hybrida
41	G1327	6552360		Nicotiana tabacum
41	G1327	8097368		Hordeum vulgare
41	G1327	8381332		Gossypium arboreum
41	G1327	9252441		Solanum tuberosum
43	G673	6062169		Lycopersicon esculentum

Figure 3E

- iguie oc		· · · · · · · · · · · · · · · · · · ·		
SEQ ID No.	GID	Genbank NID	P-value	Species
43	G673	6907081		Oryza sativa
43	G673	9205170	2.60E-28	Glycine max
43	G673	5847380	7.10E-26	Zea mays
43	G673	7614730	7.70E-26	Lotus japonicus
43	G673	9193761	3.40E-25	Medicago truncatula
43	G673	9424828	1.70E-24	Triticum aestivum
43	G673	8903196	3.40E-23	Hordeum vulgare
43	G673	6858452	2.80E-14	Sorghum bicolor
43	G673	3003284	1.50E-09	Mesembryanthemum crystallinum
45	G307	5640156	3.80E-151	Triticum aestivum
45	G307	5640154	1.00E-101	Zea mays
45	G307	6970471	1.70E-97	Oryza sativa
45	G307	7718432	4.00E-82	Medicago truncatula
45	G307	8330344	7.90E-78	Mesembryanthemum crystallinum
45	G307	5047560	1.00E-72	Gossypium hirsutum
45	G307	7588689		Glycine max
45	G307	7623983		Gossypium arboreum
45	G307	7780253		Lotus japonicus
45	G307	6733213		Lycopersicon esculentum
47	G529	1773327		Mesembryanthemum crystallinum
47	G529	8515887		Populus alba x Populus tremula
47	G529	6179980		Lilium longiflorum
47	G529	2921511		Fritillaria agrestis
47	G529	1575724		Glycine max
47	G529	466335		Lycopersicon esculentum
47	G529	1519250		Oryza sativa
47	G529	2689478		Nicotiana tabacum
47	G529	2266661		Hordeum vulgare
47	G529	1321992		Solanum tuberosum
49	G531	2921511		Fritillaria agrestis
		6179980		Lilium longiflorum
49	G531 G531	1773327		Mesembryanthemum crystallinum
49				Populus alba x Populus tremula
49	G531	8515887		Nicotiana tabacum
49	G531	2689478		
49	G531	1575724	4.90E-107	Glycine max Oryza sativa
49	G531	1519250		
49	G531	466335	1.20E-104	Lycopersicon esculentum
49	G531	1321992		Solanum tuberosum
49	G531	2266661		Hordeum vulgare
51	G214	8170933		Lycopersicon esculentum
51_	G214	9205339		Glycine max
51	G214	8577344		Zea mays
51	G214	9119112		Medicago truncatula
51	G214	7660673		Sorghum bicolor
51	G214	8213273		Oryza sativa
51	G214	3325786		Gossypium hirsutum
51	G214	9435251		Hordeum vulgare
51	G214	9411569		Triticum aestivum
51	G214	7614730		Lotus japonicus
53	G1930	7643366		Medicago truncatula
53	G1930	8329389	3.60E-47	Mesembryanthemum crystallinum
53	G1930	6069592	8.60E-47	
53	G1930	9430646	6.60E-39	Lycopersicon esculentum

Figure 3F

SEQ ID No.	GID	Genbank NID	P-value	Species
53	G1930	7722547		Lotus japonicus
53	G1930	7324245	9.20E-33	Lycopersicon pennellii
53	G1930	8902194	2.40E-31	Hordeum vulgare
53	G1930	9247126	5.90E-28	Oryza sativa
53	G1930	8749037	2.40E-27	Citrus x paradisi
53	G1930	9302986	6.40E-26	Sorghum bicolor
55	G9	7643366	5.40E-56	Medicago truncatula
55	G9	8669779		Glycine max
55	G9	8329389	1.20E-48	Mesembryanthemum crystallinum
55	G9	7412012	1.20E-41	Lycopersicon esculentum
55	G9	8902194	6.60E-36	Hordeum vulgare
55	G9	7722547		Lotus japonicus
55	G9	7324245	1.90E-32	Lycopersicon pennellii
55	G9	8749037	1.10E-31	Citrus x paradisi
55	G9	9247126	1.20E-29	Oryza sativa
55	G9	9302986	7.00E-29	Sorghum bicolor
57	G993	7643366	9.50E-59	Medicago truncatula
57	G993	8329389	8.10E-50	Mesembryanthemum crystallinum
57	G993	8669779		Glycine max
57	G993	4384549	4.20E-40	Lycopersicon esculentum
57	G993	8902194	2.00E-34	Hordeum vulgare
57	G993	7719409		Lotus japonicus
57	G993	8749037		Citrus x paradisi
57	G993	9247126		Oryza sativa
57	G993	7324245		Lycopersicon pennellii
57	G993	9302986	9.10E-27	
59	G41	5616085		Brassica napus
59	G41	5603726		Lycopersicon esculentum
59	G41	7719106		Medicago truncatula
59	G41	6667103		Glycine max
59	G41	6983854		Oryza sativa
59	G41	7324530		Lycopersicon pennellii
59	G41	8904571		Triticum aestivum
59	G41	7740143		Lotus japonicus
59	G41	7644788		Pinus taeda
59	G41	7625186		Gossypium arboreum
61	G40	5616085	7.70E-86	
61	G40	5603726		Lycopersicon esculentum
61	G40	7719106		Medicago truncatula
61	G40	6667103		Glycine max
61	G40	6983854		Oryza sativa
61	G40	8904571	3.50E-29	
61	G40	7324530	5.20E-29	
61	G40	7740143	1.40E-25	
61	G40	7644788	1.80E-20	
61	G40	7625186	5.70E-20	
63	G42	5616085	8.60E-87	
63	G42	5603726	2.20E-53	
63	G42	7719106		Medicago truncatula
63	G42	6667103	6.00E-38	
63	G42	6983854	1.10E-35	
63	G42	8904571	5.50E-31	Triticum aestivum

Figure 3G

				Ia .
SEQ ID No.	GID	Genbank NID	P-value	Species
63	G42	7324530		Lycopersicon pennellii
63	G42	7740143		Lotus japonicus
63	G42	7644788		Pinus taeda
63	G42_	7625186		Gossypium arboreum
65	G1127	6913305		Glycine max
65	G1127	9280727	5.40E-27	
65	G1127	2213533		Pisum sativum
65	G1127	7009437		Zea mays
65	G1127	7536402		Sorghum bicolor
65	G1127	7333976		Lycopersicon esculentum
65	G1127	3819186		Hordeum vulgare
65	G1127	7624850		Gossypium arboreum
65	G1127	4165182		Antirrhinum majus
65	G1127	7765939		Medicago truncatula
67	G2657	7238733		Medicago truncatula
67	G2657	6846994		Glycine max
67	G2657	7615218		Lotus japonicus
67	G2657	9445090	4.00E-41	Triticum aestivum
67	G2657	7333102		Lycopersicon esculentum
67	G2657	9252370	1.90E-27	Solanum tuberosum
67	G2657	5042437	5.90E-21	Oryza sativa
67	G2657	7536402	8.60E-20	Sorghum bicolor
67	G2657	7624850	2.20E-18	Gossypium arboreum
67	G2657	7009437	1.80E-16	Zea mays
69	G326	7410432	1.10E-37	Lycopersicon esculentum
69	G326	3618319		Oryza sativa
69	G326	7571599	4.90E-30	Medicago truncatula
69	G326	7232283		Glycine max
69	G326	7323708	6.00E-27	Lycopersicon hirsutum
69	G326	4091805	2.30E-19	Malus domestica
69	G326	6917805	6.50E-19	Lycopersicon pennellii
69	G326	3341722	2.50E-18	
69	G326	4557092	7.50E-18	
69	G326	2303680	4.70E-17	Brassica napus

mbi19 Sequence Listing.ST25 SEQUENCE LISTING

	J O P A R Y	ai-Z acqu mair ilgr dam, iech u, G amah	elinga, Pim, I im, I Luc mann uo-L	e, H ined Mars , Jo iang	eard a ha se L	,										
<120)>	Yiel	d-re	late	d ge	nes										
<130)> :	MBI-	0019													
<150 <151		60/1 1999														
<150 <151		60/1 2000														
<150 <151		Plan 2000			Modi	fica	tion	III								
<160)>	70														
<1.70) >	Pate	ntIn	ver	sion	3.0										
<210 <211		1 1084														
<212	> 1	DNA				•										
<213		Arab:	ldop	315	tnal	ıana										
<220 <221 <222 <223	.> !>	CDS (73) G26	(7:	29)												
<400 ttgg		1 gta d	ccaa	aacc	ca to	ettt	gact	t caa	aaaa	taaa	ataa	aaaa	aa (cata	aattga	60
cato	atc	gga (a at Me 1	et H	at aç is Se	gc gg er G	gg as ly Ly 5	ag ag ys Ai	ga co rg P:	ct ci ro L	ta to eu Se	ca co er P:	ro G	aa to lu Se	ca atg er Met	111
gcc Ala	gga				020	222							•			
	Gly 15	Asn	aga Arg	gaa Glu	Glu	Lys 20	Lys	gag Glu	ttg Leu	tgt Cys	tgt Cys 25	tgc	tca	act Thr	ttg Leu	159
tcg Ser 30	15 qaa	Asn tct Ser	Arg	Glu	Glu	Lys 20 qat	Lys	Glu	Leu	Cys	Cys 25 ctc	tgc Cys	tca Ser	Thr	Leu	159 207
Ser 30	gaa Glu cca	Asn	gat Asp	Glu gtg Val	tct Ser 35	Lys 20 gat Asp	ttt Phe	Glu gtc Val	tct Ser	gaa Glu 40	Cys 25 ctc Leu ctt	tgc Cys act Thr	tca Ser ggt Gly	Thr caa Gln caa	ccc Pro 45	
Ser 30 atc Ile	gaa Glu cca Pro	tct Ser tca	gat Asp tcc ser	gtg Val att Ile 50	Glu tct Ser 35 gat Asp	Lys 20 gat Asp gat Asp	ttt Phe caa Gln	gtc Val tct Ser	tct Ser tcg Ser 55	gaa Glu 40 tcg Ser	Cys 25 ctc Leu ctt Leu	tgc Cys act Thr act Thr	tca Ser ggt Gly ctt Leu	Thr caa Gln caa Gln 60 aga	ccc Pro 45 gaa Glu	207
ser 30 atc Ile aaa Lys	gaa Glu cca Pro agt Ser	tct Ser tca Ser	gat Asp tcc ser tcg ser 65	gtg Val att Ile 50 agg Arg	Glu tct Ser 35 gat Asp caa Gln	Lys 20 gat Asp gat Asp cga Arg	ttt Phe caa Gln aac Asn	gtc Val tct Ser tac Tyr 70	tct ser tcg ser 55 aga Arg	gaa Glu 40 tcg Ser ggc Gly	Cys 25 ctc Leu ctt Leu gtg Val	tgc Cys act Thr act Thr agg Arg	tca Ser ggt Gly ctt Leu caa Gln 75	Caa Gln Caa Gln 60 aga Arg	ccc Pro 45 gaa Glu ccg Pro	207 255
ser 30 atc 11e aaa Lys tgg Trp gtg Val	gaa Glu cca Pro agt Ser gga Gly	tct Ser tca Ser aac Asn aaa Lys 80	gat Asp tcc Ser tcg Ser 65 tgg Trp	gtg Val att ile 50 agg Arg gcg Ala	Glu tct Ser 35 gat Asp caa Gln gct Ala	Lys 20 gat Asp gat Asp cga Arg	ttt Phe caa Gln aac Asn att Ile 85	gtc Val tct Ser tac Tyr 70 cgt Arg	tct Ser tcg Ser 55 aga Arg gac Asp	gaa Glu 40 tcg ser ggc Gly ccg Pro	Cys 25 ctc Leu ctt Leu gtg Val aac Asn	tgc Cys act Thr act Thr agg Arg Lys 90 gcc	tca Ser ggt Gly ctt Leu caa Gln 75 gca Ala	Caa Gln Caa Gln 60 aga Arg gct Ala	ccc Pro 45 gaa Glu ccg Pro	207 255 303
atc Ile aaa Lys tgg Trp gtg Val	gaa Glu cca Pro agt Ser gga Gly tgg p5	tct Ser tca Ser aac Asn aaa Lys 80	gat Asp tcc Ser tcg Ser 65 tgg Gly gca	gtg Val att Ile 50 agg Arg gcg Ala acg Thr	tct Ser 35 gat Asp caa Gln gct Ala	Lys 20 gat Asp gat Asp cga Arg gag Glu gac Asp 100 ttt	ttt Phe caa Gln aac Asn att Ile 85 act Thr	gtc Val tct Ser tac Tyr 70 cgt Arg	tct Ser tcg Ser 55 aga Arg gac Asp	gaa Glu 40 tcg Ser ggc Gly ccg Pro	Cys 25 ctc Leu ctt Leu gtg Val aac Asn gcc Ala 105 gcc	tgc Cys act Thr act Thr agg Arg Lys 90 gcc Ala	tca Ser ggt Gly ctt Leu caa Gln 75 gca Ala tta Leu	caa Gln caa Gln 60 aga Arg gct Ala	ccc Pro 45 gaa Glu ccg Pro cgt Arg	207 255 303 351

mbil9 Sequence Listing.ST25 Pro Glu His Ile Arg Val Asn Pro Thr Gln Leu Tyr Pro Ser Pro Ala 130 135 140	
act tcc cat gat cgc att atc gtg aca cca cct agt cca cct cca cca Thr Ser His Asp Arg Ile Ile Val Thr Pro Pro Ser Pro Pro Pro 145 150 155	543
att gct cct gac ata ctt ctt gat caa tat ggc cac ttt caa tct cga Ile Ala Pro Asp Ile Leu Leu Asp Gln Tyr Gly His Phe Gln Ser Arg 160 165 170	591
agt agt gat tcc agt gcc aac ttg tcc atg aat atg ctg tct tct tcg Ser Ser Asp Ser Ser Ala Asn Leu Ser Met Asn Met Leu Ser Ser 175 180 185	639
tct tca tct ttg aat cat caa ggg cta aga cca aat ttg gag gat ggt Ser Ser Ser Leu Asn His Gln Gly Leu Arg Pro Asn Leu Glu Asp Gly 190 195 200 205	687
gaa aac gtg aag aac att agt atc cac aaa cga cga aaa taa Glu Asn Val Lys Asn Ile Ser Ile His Lys Arg Arg Lys 210 215	729
catgttaatg gcataaatat ctcttcgtcc aagttatcaa acgcattgac .ctccggcttt	789
gatcatttta ggcgcttaat ctctttacga cttcattttg gtagtcttta aagagtctat	849
ggagtggatt tagctaggaa tcaggcctta tggatgaaaa atatataaat tttgaacatg	909
actatgcaag aatgggatga agactactta gcttggaaaa cgtcctgata ggtcatgacg	969
actatateca cagaagatga eegaeggaga caacaacatg eeteaeetga tegaeegate	1029
aaatgagata atgtgttgac cggaccggtc ggatcaggtt gggtcgagta tatca	1084
<210> 2 <211> 218 <212> PRT <213> Arabidopsis thaliana	
<400> 2	
Met His Ser Gly Lys Arg Pro Leu Ser Pro Glu Ser Met Ala Gly Asn 1 10 · 15	
Arg Glu Glu Lys Lys Glu Leu Cys Cys Cys Ser Thr Leu Ser Glu Ser 20 25 30	
Asp Val Ser Asp Phe Val Ser Glu Leu Thr Gly Gln Pro Ile Pro Ser 35 40 45	
Ser Ile Asp Asp Gln Ser Ser Ser Leu Thr Leu Gln Glu Lys Ser Asn 50 60	
Ser Arg Gln Arg Asn Tyr Arg Gly Val Arg Gln Arg Pro Trp Gly Lys 75 80	
Trp Ala Ala Glu Ile Arg Asp Pro Asn Lys Ala Ala Arg Val Trp Leu 85 90 95	
Gly Thr Phe Asp Thr Ala Glu Glu Ala Ala Leu Ala Tyr Asp Lys Ala	

Ile	Arg 130	Val	Asn	Pro	Thr	Gln 135	mbi Leu	19 S Tyr	eque Pro	nce Ser	List Pro 140	ing. Ala	ST25 Thr	Ser	His	
Asp 145	Arg	Ile	Ile	Val	Thr 150	Pro	Pro	Ser	Pro	Pro 155	Pro	Pro	Ile	Ala	Pro 160	
Asp	Ile	Leu	Leu	Asp 165	Gln	Tyr	Gly	His	Phe 170	Gln	Ser	Arg	Ser	Ser 175	Asp	
Ser	Ser	Ala	Asn 180	Leu	Ser	Met	Asn	Met 185	Leu	Ser	Ser	Ser	Ser 190	Ser	Ser	
Leu	Asn	His 195	Gln	Gly	Leu	Arg	Pro 200	Asn	Leu	Glu	Asp	Gly 205	Glu	Asn	Val	
Lys	Asn 210	Ile	Ser	Ile	His	Lys 215	Arg	Arg	Lys							
<210> 3 <211> 1440 <212> DNA <213> Arabidopsis thaliana																
<213> Arabidopsis thaliana <220> <221> CDS <222> (149)(1156) <223> G38																
<400> 3 gaggaaaact cgaaaaagct acacacaaga agaagaagaa aagatacgag caagaagact 60																
			cgaaa	aaago	et ac	cacao	caaga	a aga	aagaa	agaa	aaga	atac	gag (caaga	aagact	60
gag	gaaa	act o	_												aagact gtcccc	60 120
gag	gaaaa	act o	gcgat	ttat	c aa	actc	gaagg gag a		gagad gca g	ettt gtt (gati tat q Tyr	tttca	aaa t	tttc agt (gteece gga	
gagg	gaaaa cacga agati	act of an act	gcgatgttgt	ttat ttct	caa	actco	gaagg gag a l :	g aag atg g Met A	gagad gca q Ala v	ettt gtt (/al :	gati tat q Tyr ;	gat (Asp (aaa t Cag a Gln S	agt of tot	gtcccc gga Gly aga	120
gagg aaac tata gat Asp	agati agati aga aga aga aga	act of ac	gcgat gttgt aga Arg	aca Thr	caa Gln	actco gaagg att Ile 15	gaagg gag a gat Asp	g aag atg g Met <i>l</i> l aca	gagad gca q Ala i tcg ser	agg Arg	gati	gat (Asp (S	aaa taaa aaa Lys	agt (Ser (tct Ser tgg	gtcccc gga gly aga Arg	120 172
gagg aaaa tata gat Asp agt ser 25	agati agati aga aga aga arg 10 aga Arg	aaa gaac gaac Asn ggt Gly	gcgat gttgt aga Arg gac Asp	aca Thr ggt Gly	caa Gln act Thr 30	actco gaagg att Ile 15 act Thr	gaagg gag a gat Asp gtg Val	atg of the second secon	gagad gca g Ala v tcg Ser gag Glu	agg Arg aga Arg 35	gati tat of Tyr i aaa Lys 20 tta Leu	gat (Asp (5) agg Arg aag Lys	aaa (31n S aaa Lys aga Arg	tct Ser (tct Ser tgg Trp	gtcccc gga gga aga Arg aaa Lys 40 gta	120 172 220
gagg aaaa tata gat Asp agt Ser 25 gag Glu	aga agati aga aga aga Arg 10 aga Arg tat Tyr	aaa gaac Asn ggt Gly	aga Arg gac Asp gag Glu	aca Thr ggt Gly acc Thr 45	caa Gln act Thr 30 gta Val	actco gaagg att Ile 15 act Thr gaa Glu	gaaggag a la l	atg of Ala	gagad gca d Ala v tcg Ser gag Glu tct Ser 50	agg Arg aga Arg 35 acc Thr	gati tat of Tyr ; aaaa Lys 20 tta Leu aag Lys	gat (Asp (Solution)	aaa taaa taaa Lys aga Arg	tttcq agt (Ser (tct Ser tgg Trp aaa Lys 55	gtcccc gga gga aga Arg aaa Lys 40 gta Val	120 172 220 268
gagggagaaaaatatatatatatatatatatatatatat	agat Arg 10 aga Arg tat Tyr	aac Asn aaa Lys agc	gagagagagagagagagagagagagagagagagagaga	aca Thr ggt gly acc Thr 45 tcg Ser	caa Gin act Thr 30 gta Val aag	att Ile 15 act Thr gaa ag Lys	gaaggagagagagagagagagagagagagagagagaga	atg of the state o	ggaga ggaga tcg Ser gag Glu tct Ser 50 atg Met	agg Arg aga Arg 35 acc Thr	gatt tat () tat () aaaa Lys 20 tta Leu aag Lys Gly	gat (Asp (Asp (Asp (Asp (Asp (Asp (Asp (Asp	aaaa ligaa aaaa Lys aaga Arg aagg Arg gga agg aagg aagg aag	tcttcq Ger (tct Ser tgg Trp aaaa Lys 55 gga Gly	gtcccc gga gga aga Arg aaa Lys 40 gta Val cca Pro	120 172 220 268 316
gagggagagatatatatatatatatatatatatatatat	aga Arg 10 aga Arg tat Tyr gcg Ala aat Asn	aac Asn aaa Lys agc Ser 75	ggggat gttgt aga Arg gac Asp gag Glu ggg Gly 60 cga Arg	aca Thr ggt Gly acc Thr 45 tcg Ser tgt Cys	caa Gin act Thr 30 gta Val aag Lys agt	att Ile 15 act Thr gaa agg Lys	gaaggag a light of the second	gaag Met 1 aca Thr gct Ala gtt Val	gagagagagagagagagagagagagagagagagagaga	agg Arg acc Thr aaaa Lys agg Arg	gatt tat () tat () aaaa Lys 20 tta Leu aag Lys Gly caa Gln	aagg Arg aaaa Lys aagg Arg Arg aagg Arg aagg Arg aagg Arg aagg Arg agg Arg agg agg	aga Arg agg Arg agg Arg agg Arg agg Arg agg agg Arg agg agg agg agg agg agg agg agg agg a	tcttcq Ser (tctt Ser tgg Trp aaaa Lys 55 gga Gly tgg Trp	gtcccc gga gga aga Arg aaa Lys 40 gta Val cca Pro ggt Gly	120 172 220 268 316
gagggagagatatatatatatatatatatatatatatat	gaaaacacacacacacacacacacacacacacacacaca	act of case of	aga Arg gac Asp gac Asp Glu gag Gly 60 cga Arg	aca Thr ggt Gly acc Thr 45 tcg Ser tgt Cys	cc aa Gg gg caa Gln act Thr 30 gta Val aag Lys agt Ser atc Ile	atterion atterior actor	gaaggaaggaggaggaggaggaggaggaggaggaggagg	get Ala gtt Val tgt Cys 65 gga Gly	gagagagagagagagagagagagagagagagagagaga	agg Arg aga Arg 35 acc Thr aaa Lys agg Arg	gattat (fyr)	aagg Lys aagg Lys aagg Lys aagg Ser agg	aga Arg gga Arg agg Arg tat tat	tctc ser (tct Ser tgg Trp aaaa Lys 55 gga Gly tgg Trp	gtcccc gga gily aga Arg aaa Lys 40 gta Val cca Pro ggt Gly tgg Trp	120 172 220 268 316 364
gagggagatatatatatatatatatatatatatatatat	agaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	aac Asn ggt Gly aac Asn aga Lys ggtt Val act Thr	gcgat aga gac Asp gac Asp gag Glu ggg Gly 60 cga Arg	aca Thr ggt Gly acc Thr 45 tcg Ser tgt Cys	caa Gln act Thr 30 gta Val aag Lys agt Ser atc Ile	attering gaa gguagg	gaaggaaggaggaggaggaggaggaggagggaaggagggagggagggagggagggagggagggagggagggagggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggag	gaaquatg Gaaca Thr Ala Stt Val tgt Cys 65 gga Gly Cct Pro	gagaga gca gca Ala v tcg Ser gag Glu tct Ser 50 atg Met Val aat Asn gct Ala 130	aga Arg aga Arg acc Thr aaa Lys agg Arg cga Arg cga Arg cga CT CGA	gatt tat () Tyr) aaaa Lys 20 tta Lys Lys Gly caa Gln ggt Gly tct Ser	aagg Lys aaag Lys aagg Arg aag Cys aaag Cys aaag Cys aaaa Cys aagg Arg 85 agc Ser	aga Arg agg Arg agg Arg att Ile agg Arg	tct con tct co	gtcccc gga gga gly aga Arg aaa Lys 40 gta Val cca Pro ggt Gly tgg Trp gag Glu 120 cgg	120 172 220 268 316 364 412

PCT/US00/31414 WO 01/35725

mbi19 Sequence Listing.ST25

									•			•				
tct Ser	gat Asp	gcg Ala	tct Ser 140	gag Glu	gtt Val	acg Thr	agt Ser	acc Thr 145	tca Ser	agt Ser	cag Gln	tct Ser	gag Glu 150	gtg Val	tgt Cys	604
act Thr	gtt Val	gag Glu 155	act Thr	cct Pro	ggt Gly	tgt Cys	gtt Val 160	cat His	gtg Val	aaa Lys	aca Thr	gag Glu 165	gat Asp	cca Pro	gat Asp	652
tgt Cys	gaa Glu 170	tct Ser	aaa Lys	ccc Pro	ttc Phe	tcc Ser 175	ggt Gly	gga Gly	gtg Val	gag Glu	ccg Pro 180	atg Met	tat Tyr	tgt Cys	ctg Leu	700
gag Glu 185	aat Asn	ggt Gly	gcg Ala	gaa Glu	gag Glu 190	atg Met	aag Lys	aga Arg	ggt Gly	gtt Val 195	aaa Lys	gcg Ala	gat Asp	aag Lys	cat His 200	748
tgg Trp	ctg Leu	agc Ser	gag Glu	ttt Phe 205	gaa Glu	cat His	aac Asn	tat Tyr	tgg Trp 210	agt Ser	gat Asp	att Ile	ctg Leu	aaa Lys 215	gag Glu	796
aaa Lys	gag Glu	aaa Lys	cag Gln 220	aag Lys	gag Glu	caa Gln	999 Gly	att Ile 225	gta Val	gaa Glu	acc Thr	tgt Cys	cag Gln 230	caa Gln	caa Gln	844
cag Gln	cag Gln	gat Asp 235	tcg Ser	cta Leu	tct Ser	gtt Val	gca Ala 240	gac Asp	tat Tyr	ggt Gly	tgg Trp	ccc Pro 245	aat Asn	gat Asp	gtg Val	892
gat Asp	cag Gln 250	agt Ser	cac His	ttg Leu	gat Asp	tct Ser 255	tca Ser	gac Asp	atg Met	ttt Phe	gat Asp 260	gtc Val	gat Asp	gag Glu	ctt Leu	940
cta Leu 265	cgt Arg	gac Asp	cta Leu	aat Asn	ggc Gly 270	gac Asp	gat Asp	gtg Val	ttt Phe	gca Ala 275	ggc Gly	tta Leu	aat Asn	cag Gln	gac Asp 280	988
cgg Arg	tac Tyr	ccg Pro	61y 999	aac Asn 285	agt Ser	gtt Val	gcc Ala	aac Asn	ggt Gly 290	tca Ser	tac Tyr	agg Arg	ccc Pro	gag Glu 295	agt Ser	1036
caa Gln	caa Gln	agt Ser	ggt Gly 300	ttt Phe	gat Asp	ccg Pro	cta Leu	caa Gln 305	agc Ser	ctc Leu	aac Asn	tac Tyr	gga Gly 310	ata Ile	cct Pro	1084
ccg Pro	ttt Phe	cag Gln 315	ctc Leu	gag Glu	gga Gly	aag Lys	gat Asp 320	ggt Gly	aat Asn	gga Gly	ttc Phe	ttc Phe 325	gac Asp	gac Asp	ttg Leu	1132
					gag Glu		taa	acaa	aaaca	aat a	atgaa	igctt	t tt	ggat	ttga	1186
tatt	tgcc	ett a	atco	caca	a cç	acto	sttga	a tto	tcta	atcc	gagt	ttta	gt g	jatat	agaga	1246
acta	caga	ac a	cgtt	ttt	c tt	gtta	taaa	a ggt	gaad	tgt	atat	atco	gaa a	acagt	gatat	1306
															agata	
tgtt	ttat	gt t	ttgt	aaca	aa ca	ıggaa	tgaa	a taa	ataca	cac	ttgt	gaag	jct t	ttaa	aaaaa	1426
aaaa	aaaa	aa a	aaa													1440

<210> 4 <211> 335 <212> PRT <213> Arabidopsis thaliana

<400> 4

mbil9 Sequence Listing.ST25
Thr Ser Arg Lys Arg Lys Ser Arg Ser Arg Gly Asp Gly Thr Thr Val
20 25 30

Ala Glu Arg Leu Lys Arg Trp Lys Glu Tyr Asn Glu Thr Val Glu Glu 40 45

Val Ser Thr Lys Lys Arg Lys Val Pro Ala Lys Gly Ser Lys Lys Gly 50 55 60

Cys Met Lys Gly Lys Gly Gly Pro Glu Asn Ser Arg Cys Ser Phe Arg 65 70 75 80

Gly Val Arg Gln Arg Ile Trp Gly Lys Trp Val Ala Glu Ile Arg Glu 85 90 95

Pro Asn Arg Gly Ser Arg Leu Trp Leu Gly Thr Phe Pro Thr Ala Gln 100 105 110

Glu Ala Ala Ser Ala Tyr Asp Glu Ala Ala Lys Ala Met Tyr Gly Pro 115 120 125

Leu Ala Arg Leu Asn Phe Pro Arg Ser Asp Ala Ser Glu Val Thr Ser 130 140

Thr Ser Ser Gln Ser Glu Val Cys Thr Val Glu Thr Pro Gly Cys Val 145 150 150 155

His Val Lys Thr Glu Asp Pro Asp Cys Glu Ser Lys Pro Phe Ser Gly
165 170 175

Gly Val Glu Pro Met Tyr Cys Leu Glu Asn Gly Ala Glu Glu Met Lys 180 185 190

Arg Gly Val Lys Ala Asp Lys His Trp Leu Ser Glu Phe Glu His Asn 195 200 205

Tyr Trp Ser Asp Ile Leu Lys Glu Lys Glu Lys Gln Lys Glu Gln Gly 210 215 220

Ile Val Glu Thr Cys Gln Gln Gln Gln Gln Asp Ser Leu Ser Val Ala 225 230 235 240

Asp Tyr Gly Trp Pro Asn Asp Val Asp Gln Ser His Leu Asp Ser Ser 245 250 255

Asp Met Phe Asp Val Asp Glu Leu Leu Arg Asp Leu Asn Gly Asp Asp 260 265 270

Val Phe Ala Gly Leu Asn Gln Asp Arg Tyr Pro Gly Asn Ser Val Ala 275 280 285

Asn Gly Ser Tyr Arg Pro Glu Ser Gln Gln Ser Gly Phe Asp Pro Leu 290 295 300

Gln Ser Leu Asn Tyr Gly Ile Pro Pro Phe Gln Leu Glu Gly Lys Asp 305 310 315 320

mbi19 Sequence Listing.ST25

Gly Asn Gly Phe Phe Asp Asp Leu Ser Tyr Leu Asp Leu Glu Asn 325 330 335

<210> <211> 909 <212> DNA Arabidopsis thaliana <220> <221> CDS <222> (38) . . (643) <400> 5 ctcctgtctt gtctaaagaa aaaagagaga ggaagaa atg gag act ttt gag gaa Met Glu Thr Phe Glu Glu 103 age tet gat ttg gat gtt ata eag aaa cat eta ttt gaa gae ttg atg Ser Ser Asp Leu Asp Val Ile Gln Lys His Leu Phe Glu Asp Leu Met atc cct gat ggt ttc att gaa gat ttt gtc ttt gat gat act gct ttt Ile Pro Asp Gly Phe Ile Glu Asp Phe Val Phe Asp Asp Thr Ala Phe 151 gtc tcc gga ctc tgg tct cta gaa ccc ttt aac cca gtt ccg aaa ctg Val Ser Gly Leu Trp Ser Leu Glu Pro Phe Asn Pro Val Pro Lys Leu gaa cct agt tca cct gtt ctt gat cca gat tcc tat gtc caa gag att Glu Pro Ser Ser Pro Val Leu Asp Pro Asp Ser Tyr Val Gln Glu Ile 247 ctg caa atg gaa gca gaa tca tca tca tca tca tca aca aca acg tca Leu Gln Met Glu Ala Glu Ser Ser Ser Ser Ser Ser Thr Thr Thr Ser 295 cct gag gtt gag act gtc tca aac cgg aaa aaa aca aag agg ttt gaa Pro Glu Val Glu Thr Val Ser Asn Arg Lys Lys Thr Lys Arg Phe Glu 343 gaa acg aga cat tac aga ggc gtg aga agg agg cca tgg ggg aaa ttt Glu Thr Arg His Tyr Arg Gly Val Arg Arg Arg Pro Trp Gly Lys Phe 391 110 gca gca gag att cga gat ccg gca aag aaa gga tcc agg att tgg tta Ala Ala Glu Ile Arg Asp Pro Ala Lys Lys Gly Ser Arg Ile Trp Leu 120 125 130 439 ggc act ttt gag agt gat att gat gct gca agg gct tac gac tat gca Gly Thr Phe Glu Ser Asp Ile Asp Ala Ala Arg Ala Tyr Asp Tyr Ala 487 gct ttt aag ctc agg gga aga aaa gct gtt ctc aac ttt cct ttg gat Ala Phe Lys Leu Arg Gly Arg Lys Ala Val Leu Asn Phe Pro Leu Asp 535 gcc gga aag tat gat gct ccg gtc aat tca tgc cga aaa agg agg aga Ala Gly Lys Tyr Asp Ala Pro Val Asn Ser Cys Arg Lys Arg Arg Arg Arg acc gat gta cca cag cct caa gga aca aca aca agt act tca tca tcg
Thr Asp Val Pro Gln Pro Gln Gly Thr Thr Thr Ser Thr Ser Ser Ser 631 tca tca aac taa tgggggaata gtgatgttta attagtatat ataggttaat 683 Ser Ser Asn atcttaagta tgtgaagcat catgtataga gccaagaacc tgttagacta gtgtactgaa

803

863 909

mbi19 Sequence Listing.ST25 aagaactett geaaaatatg taetaaagag tteetgtaae aatggaaett etgegtttte tcttgtctta aagagcttaa ggttctagaa acaaagttct tgtcctttcg gtttaaaaaaa <210> 6 <211> 201 <212> PRT <213> Arabidopsis thaliana <400> 6 Met Glu Thr Phe Glu Glu Ser Ser Asp Leu Asp Val Ile Gln Lys His 1 10 15 Leu Phe Glu Asp Leu Met Ile Pro Asp Gly Phe Ile Glu Asp Phe Val 20 30 Phe Asp Asp Thr Ala Phe Val Ser Gly Leu Trp Ser Leu Glu Pro Phe Asn Pro Val Pro Lys Leu Glu Pro Ser Ser Pro Val Leu Asp Pro Asp 50 55 60 Ser Tyr Val Gln Glu Ile Leu Gln Met Glu Ala Glu Ser Ser Ser Ser Ser Ser Thr Thr Thr Ser Pro Glu Val Glu Thr Val Ser Asn Arg Lys Lys Thr Lys Arg Phe Glu Glu Thr Arg His Tyr Arg Gly Val Arg Arg 100 105 110 Arg Pro Trp Gly Lys Phe Ala Ala Glu Ile Arg Asp Pro Ala Lys Lys $115 \ \ 120 \ \ \cdot \ 125$ Gly Ser Arg Ile Trp Leu Gly Thr Phe Glu Ser Asp Ile Asp Ala Ala 130 135 140 Arg Ala Tyr Asp Tyr Ala Ala Phe Lys Leu Arg Gly Arg Lys Ala Val 145 150 155 160 Leu Asn Phe Pro Leu Asp Ala Gly Lys Tyr Asp Ala Pro Val Asn Ser Cys Arg Lys Arg Arg Arg Thr Asp Val Pro Gln Pro Gln Gly Thr Thr Thr Ser Thr Ser Ser Ser Ser Ser Asn <210> 7 <211> 1107 <212> DNA <213> Arabidopsis thaliana

Page 7

<220>

<221> CDS <222> (16)..(930) <223> G207

mbil9 Sequence Listing.ST25

<pre><400> 7 aaaaagatctg tttca atg gcg gat cgt gtt aaa ggt cca tgg agt caa gaa</pre>													51			
gaa Glu	gat Asp	gag Glu 15	cag Gln	cta Leu	cga Arg	agg Arg	atg Met 20	gtt Val	gag Glu	aaa Lys	tac Tyr	gga Gly 25	ccg Pro	agg Arg	aat Asn	99
tgg Trp	tct Ser 30	gcg Ala	att Ile	agc Ser	aaa Lys	tcg Ser 35	att Ile	cca Pro	ggt Gly	cga Arg	tct Ser 40	ggt Gly	aaa Lys	tcg Ser	tgt Сув	147
aga Arg 45	tta Leu	cgt Arg	tgg Trp	tgt Cys	aat Asn 50	cag Gln	tta Leu	tct Ser	ccg Pro	gag Glu 55	gtt Val	gag Glu	cat His	cgt Arg	cct Pro 60	195
ttc Phe	tcg Ser	ccg Pro	gag Glu	gaa Glu 65	gat Asp	gag Glu	act Thr	att Ile	gta Val 70	acc Thr	gcc Ala	cgt Arg	gct Ala	cag Gln 75	ttt Phe	243
ggt Gly	aac Asn	aag Lys	tgg Trp 80	gcg Ala	acg Thr	att Ile	gct Ala	cgt Arg 85	ctt Leu	ctt Leu	aac Asn	ggt Gly	cgt Arg 90	acg Thr	gat Asp	291
aac Asn	gcc Ala	gtt Val 95	aaa Lys	aat Asn	cac His	tgg Trp	aac Asn 100	tct Ser	acg Thr	ctt Leu	aag Lys	agg Arg 105	aaa Lys	tgc Cys	agc Ser	339
gga Gly	ggt Gly 110	gtg Val	gcg Ala	gtt Val	acg Thr	acg Thr 115	gtg Val	acg Thr	gag Glu	acg Thr	gag Glu 120	gaa Glu	gat Asp	cag Gln	gat Asp	387
cgg Arg 125	ccg Pro	aag Lys	aag Lys	agg Arg	aga Arg 130	tct Ser	gtt Val	agc Ser	ttt Phe	gat Asp 135	cct Pro	gct Ala	ttt Phe	gct Ala	ccg Pro 140	435
gtg Val	gat Asp	act Thr	gga Gly	ttg Leu 145	tac Tyr	atg Met	agt Ser	cct Pro	gag Glu 150	agt Ser	cct Pro	aac Asn	gga Gly	atc Ile 155	gat Asp	483
gtt Val	agt Ser	gat Asp	tct Ser 160	agc Ser	acg Thr	att Ile	ccg Pro	tca Ser 165	ccg Pro	tcg Ser	tct Ser	cct Pro	gtt Val 170	gct Ala	cag Gln	531
ctg Leu	ttt Phe	aaa Lys 175	cca Pro	atg Met	ccg Pro	att Ile	tcc Ser 180	ggc Gly	ggt Gly	ttt Phe	acg Thr	gtg Val 185	gtt Val	ccg Pro	cag Gln	579
ccg Pro	tta Leu 190	ccg Pro	gtt Val	gaa Glu	atg Met	tct Ser 195	tcg Ser	tct Ser	tcg Ser	gag Glu	gat Asp 200	cca Pro	cct Pro	act Thr	tcg Ser	627
ttg Leu 205	agt Ser	ttg Leu	tca Ser	cta Leu	cct Pro 210	gga Gly	gct Ala	gag Glu	aac Asn	acg Thr 215	agt Ser	tcg Ser	agc Ser	cat His	aac Asn 220	675
aat Asn	aac Asn	aac Asn	aac Asn	gcg Ala 225	ttg Leu	atg Met	ttt Phe	ccg Pro	aga Arg 230	ttt Phe	gag Glu	agt Ser	cag Gln	atg Met 235	aag Lys	723
att Ile	aat Asn	gta Val	gag Glu 240	gag Glu	aga Arg	gga Gly	gga Gly	gga Gly 245	gga Gly	gaa Glu	gga Gly	cgt Arg	aga Arg 250	ggt Gly	gag Glu	771
					cag Gln											819
atg Met	gcg Ala 270	gaa Glu	atg Met	cag Gln	aaa Lys	aca Thr 275	agt Ser	ggt Gly	gga Gly	ttc Phe	gtc Val 280	gtc Val	gga Gly	ggt Gly	tta Leu	867

mbil9 Sequence Listing.ST25 tac gaa tcc ggc ggc aat ggt ggt ttt agg gat tgt gga gta ata a													
cct aag gtt gag tag ttttggttta gggttaaaac ttgaatcgat tggggatttt Pro Lys Val Glu													
caagagcatt catttttggg gtttatggta aaattaaaaa caaaaacaaa atgtacagag													
gaattaaaat ttctatggaa taatcttaaa tctcaaatat ttgttacttg ttttggtgat													
tcataaccaa aatcaaa													
<210> 8 <211> 304 <212> PRT <213> Arabidopsis thaliana													
<400> 8													
Met Ala Asp Arg Val Lys Gly Pro Trp Ser Gln Glu Glu Asp Glu Gln 1 10 15													
Leu Arg Arg Met Val Glu Lys Tyr Gly Pro Arg Asn Trp Ser Ala Ile 20 25 30													
Ser Lys Ser Ile Pro Gly Arg Ser Gly Lys Ser Cys Arg Leu Arg Trp 35 . 40 45													
Cys Asn Gln Leu Ser Pro Glu Val Glu His Arg Pro Phe Ser Pro Glu 50 60													
Glu Asp Glu Thr Ile Val Thr Ala Arg Ala Gln Phe Gly Asn Lys Trp 65 70 75 80													
Ala Thr Ile Ala Arg Leu Leu Asn Gly Arg Thr Asp Asn Ala Val Lys 85 90 95													
Asn His Trp Asn Ser Thr Leu Lys Arg Lys Cys Ser Gly Gly Val Ala 100 105 110													
Val Thr Thr Val Thr Glu Thr Glu Glu Asp Gln Asp Arg Pro Lys Lys 115 120 125													
Arg Arg Ser Val Ser Phe Asp Pro Ala Phe Ala Pro Val Asp Thr Gly 130 135 140													
Leu Tyr Met Ser Pro Glu Ser Pro Asn Gly Ile Asp Val Ser Asp Ser 145 150 155 160													
Ser Thr Ile Pro Ser Pro Ser Ser Pro Val Ala Gln Leu Phe Lys Pro 165 170 175													
Met Pro Ile Ser Gly Gly Phe Thr Val Val Pro Gln Pro Leu Pro Val 180 185 190													
Glu Met Ser Ser Ser Ser Glu Asp Pro Pro Thr Ser Leu Ser Leu Ser 195 200 205													
Leu Pro Gly Ala Glu Asn Thr Ser Ser Ser His Asn Asn Asn Asn													

mbil9 Sequence Listing.ST25

210 Ala Leu Met Phe Pro Arg Phe Glu Ser Gln Met Lys Ile Asn Val Glu Glu Arg Gly Gly Gly Glu Gly Arg Arg Gly Glu Phe Met Thr Val 245 250 255 Val Gln Glu Met Ile Lys Ala Glu Val Arg Ser Tyr Met Ala Glu Met Gln Lys Thr Ser Gly Gly Phe Val Val Gly Gly Leu Tyr Glu Ser Gly Gly Asn Gly Gly Phe Arg Asp Cys Gly Val Ile Thr Pro Lys Val Glu 290 295 300 <210> <211> 1046 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (46)..(867) G241 <400> 9 gaaaaacatt tcaacttctt ttatcagcaa tcacaaatca aagag atg gga aga gct 57 Met Gly Arg Ala cca tgc tgt gag aag atg ggg ttg aag aga gga cca tgg aca cct gaa Pro Cys Cys Glu Lys Met Gly Leu Lys Arg Gly Pro Trp Thr Pro Glu gaa gat caa atc ttg gtc tct ttt atc ctc aac cat gga cat agt aac Glu Asp Gln Ile Leu Val Ser Phe Ile Leu Asn His Gly His Ser Asn 153 tgg cga gcc ctc cct aag caa gct ggt ctt ttg aga tgt gga aaa agc Trp Arg Ala Leu Pro Lys Gln Ala Gly Leu Leu Arg Cys Gly Lys Ser 201 tgt aga ctt agg tgg atg aac tat tta aag cct gat att aaa cgt ggc Cys Arg Leu Arg Trp Met Asn Tyr Leu Lys Pro Asp Ile Lys Arg Gly 249 aat ttc acc aaa gaa gag gaa gat gct atc atc agc tta cac caa ata Asn Phe Thr Lys Glu Glu Glu Asp Ala Ile Ile Ser Leu His Gln Ile 297 ctt ggc aat aga tgg tca gcg att gca gca aaa ctg cct gga aga acc Leu Gly Asn Arg Trp Ser Ala Ile Ala Ala Lys Leu Pro Gly Arg Thr 85 90 95 100 345 gat aac gag atc aag aac gta tgg cac act cac ttg aag aag aga ctc Asp Asn Glu Ile Lys Asn Val Trp His Thr His Leu Lys Lys Arg Leu 393 110 gaa gat tat caa cca gct aaa cct aag acc agc aac aaa aag aag ggt 441 Glu Asp Tyr Gln Pro Ala Lys Pro Lys Thr Ser Asn Lys Lys Gly act aaa cca aaa tct gaa tcc gta ata acg agc tcg aac agt act aga Thr Lys Pro Lys Ser Glu Ser Val Ile Thr Ser Ser Asn Ser Thr Arg

	~~~	<b>.</b>	~~~	cta	gca	gat		19 S						age	tta	537
Ser	Glu 150	Ser	Glu	Leu	Ala	Asp 155	Ser	Ser	Asn	Pro	Ser 160	Gly	Glu	Ser	Leu	33.
ttt Phe 165	tcg Ser	aca Thr	tcg Ser	cct Pro	tcg Ser 170	aca Thr	agt Ser	gag Glu	gtt Val	tct Ser 175	tcg Ser	atg Met	aca Thr	ctc Leu	ata Ile 180	585
agc Ser	cac His	gac Asp	ggc Gly	tat Tyr 185	agc Ser	aac Asn	gag Glu	att Ile	aat Asn 190	atg Met	gat Asp	aac Asn	aaa Lys	ccg Pro 195	gga Gly	633
gat Asp	atc Ile	agt Ser	act Thr 200	atc Ile	gat Asp	caa Gln	gaa Glu	tgt Cys 205	gtt Val	tct Ser	ttc Phe	gaa Glu	act Thr 210	ttt Phe	ggt Gly	681
gcg Ala	gat Asp	atc Ile 215	gat Asp	gaa Glu	agc Ser	ttc Phe	tgg Trp 220	aaa Lys	gag Glu	aca Thr	ctg Leu	tat Tyr 225	agc Ser	caa Gln	gat Asp	729
gaa Glu	cac His 230	aac Asn	tac Tyr	gta Val	tcg Ser	aat Asn 235	gac Asp	cta Leu	gaa Glu	gtc Val	gct Ala 240	ggt Gly	tta Leu	gtt Val	gag Glu	777
ata Ile 245	caa Gln	caa Gln	gag Glu	ttt Phe	caa Gln 250	aac Asn	ttg Leu	ggc Gly	tcc Ser	gct Ala 255	aat Asn	aat Asn	gag Glu	atg Met	att Ile 260	825
ttt Phe	gac Asp	agt Ser	gag Glu	atg Met 265	gaa Glu	ctt Leu	ctg Leu	gtt Val	cga Arg 270	tgt Cys	att Ile	ggc Gly	tag			867
aaccggcggg gaacaagatc tcttagccgg gctctagtta acatgtttga ggagtaaagt													927			
gaaatggtgc aaattagtta aggctaagaa attcaaaagc ttttgtttac cgagaaaaaa 987																
acad	cacto	cta a	actct	tgat	g to	atgt	agt	agt	gtat	taa	ttag	gaggo	tg (	gtti	tcaa	1046
<21 <21	acacactcta actcttgatg tgatgtagtt agtgtattaa ttagaggctg cgttttcaa 1046  <210> 10 <211> 273 <212> PRT <213> Arabidopsis thaliana															
<400	)> :	10														
Met 1	Gly	Arg	Ala	Pro 5	Сув	Сув	Glu	Lys	Met 10	Gly	Leu	Lys	Arg	Gly 15	Pro	
Trp	Thr	Pro														
			Glu 20	Glu	Asp	Gln	Ile	Leu 25	Val	Ser	Phe	Ile	Leu 30	Asn	His	
Gly	His		20		Asp Arg			25					30			
		Ser 35	20 Asn	Trp		Ala	Leu 40	25 Pro	Lys	Gln	Ala	Gly 45	30 Leu	Leu	Arg	
Сув	Gly 50	Ser 35 Lys	20 Asn Ser	Trp Cys	Arg	Ala Leu 55	Leu 40 Arg	25 Pro Trp	Lys Met	Gln Asn	Ala Tyr 60	Gly 45 Leu	30 Leu Lys	Leu Pro	Arg Asp	
Cys Ile 65	Gly 50 Lys	Ser 35 Lys Arg	20 Asn Ser Gly	Trp Cys Asn	Arg Arg	Ala Leu 55	Leu 40 Arg Lys	25 Pro Trp Glu	Lys Met Glu	Gln Asn Glu 75	Ala Tyr 60 Asp	Gly 45 Leu Ala	Leu Lys	Leu Pro Ile	Arg Asp Ser 80	
Cys Ile 65 Leu	Gly 50 Lys His	Ser 35 Lys Arg Gln	20 Asn Ser Gly	Trp Cys Asn Leu 85	Arg Arg Phe	Ala Leu 55 Thr	Leu 40 Arg Lys	Pro Trp Glu Trp	Lys Met Glu Ser 90	Gln Asn Glu 75 Ala	Ala Tyr 60 Asp	Gly 45 Leu Ala	Leu Lys Ile	Leu Pro Ile Lys 95	Arg Asp Ser 80	

Lys Lys Arg Leu Glu Asp Tyr Gln Pro Ala Lys Pro Lys Thr Ser Asn

PCT/US00/31414 WO 01/35725

mbi19 Sequence Listing.ST25 115

Lys Lys Lys Gly Thr Lys Pro Lys Ser Glu Ser Val Ile Thr Ser Ser

Asn Ser Thr Arg Ser Glu Ser Glu Leu Ala Asp Ser Ser Asn Pro Ser

Gly Glu Ser Leu Phe Ser Thr Ser Pro Ser Thr Ser Glu Val Ser Ser

Met Thr Leu Ile Ser His Asp Gly Tyr Ser Asn Glu Ile Asn Met Asp

Asn Lys Pro Gly Asp Ile Ser Thr Ile Asp Gln Glu Cys Val Ser Phe 195 200 205

Glu Thr Phe Gly Ala Asp Ile Asp Glu Ser Phe Trp Lys Glu Thr Leu

Tyr Ser Gln Asp Glu His Asn Tyr Val Ser Asn Asp Leu Glu Val Ala

Gly Leu Val Glu Ile Gln Gln Glu Phe Gln Asn Leu Gly Ser Ala Asn

Asn Glu Met Ile Phe Asp Ser Glu Met Glu Leu Leu Val Arg Cys Ile 260 265 270

Gly

<210> 11 <211> 1391 <212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS <222> (15)..(923) <223> G254

cgatttcgag ctct atg gtg tcc gta aac cct aga cct aag ggt ttt cca Met Val Ser Val Asn Pro Arg Pro Lys Gly Phe Pro

gtt ttc gat tcc tcg aat atg agt tta cca agc tcc gat gga ttt ggt Val Phe Asp Ser Ser Asn Met Ser Leu Pro Ser Ser Asp Gly Phe Gly 15 20 25 98

tcg att ccg gcc acg gga ccg acc agt acg gtg tcg ttt tct gag gat Ser Ile Pro Ala Thr Gly Arg Thr Ser Thr Val Ser Phe Ser Glu Asp 30 35 40

ccg acg acg aag att cgg aag ccg tac aca atc aag aag tcg aga gag Pro Thr Thr Lys Ile Arg Lys Pro Tyr Thr Ile Lys Lys Ser Arg Glu 194

aat tgg aca gat caa gag cac gat aaa ttt cta gaa gct ctt cac tta Asn Trp Thr Asp Gln Glu His Asp Lys Phe Leu Glu Ala Leu His Leu 70

							mhi	10 5	emie	nce '	List	i na	ST25			
ttc Phe	gat Asp	agg Arg	gat Asp 80	tgg Trp	aag Lys	aaa Lys	ata	gaa	gcc	ttt	gtt	gga	tca	aaa Lys	aca Thr	290
gta Val	gtt Val	cag Gln 95	ata Ile	cga Arg	agc Ser	cac His	gct Ala 100	cag Gln	aaa Lys	tac Tyr	ttt Phe	ctc Leu 105	aaa Lys	gtt Val	cag Gln	338
aag Lys	agt Ser 110	ggt Gly	gct Ala	aac Asn	gaa Glu	cat His 115	ctt Leu	cca Pro	ctt Leu	cct Pro	cga Arg 120	cct Pro	aag Lys	agg Arg	aaa Lys	386
gcg Ala 125	agt Ser	cat His	cct Pro	tat Tyr	cct Pro 130	ata Ile	aag Lys	gct Ala	cct Pro	aaa Lys 135	aat Asn	gtt Val	gct Ala	tat Tyr	acc Thr 140	434
tct Ser	ctc Leu	ccg Pro	tct Ser	tcg Ser 145	agt Ser	aca Thr	tta Leu	ccg Pro	ttg Leu 150	ctt Leu	gag Glu	cct Pro	ggt Gly	tat Tyr 155	ttg Leu	482
tat Tyr	agc Ser	tct Ser	gat Asp 160	tcg Ser	aag Lys	tca Ser	ttg Leu	atg Met 165	gga Gly	aac Asn	cag Gln	gct Ala	gtt Val 170	tgt Cys	gca Ala	530
tct Ser	acc Thr	tct Ser 175	tct Ser	tcg Ser	tgg Trp	aat Asn	cat His 180	gaa Glu	tcg Ser	aca Thr	aat Asn	ctg Leu 185	cca Pro	aaa Lys	ccg Pro	578
gtg Val	att Ile 190	gaa Glu	gag Glu	gaa Glu	ccg Pro	gga Gly 195	gtc Val	tcg Ser	gcc Ala	acg Thr	gct Ala 200	cct Pro	ctc Leu	cca Pro	aat Asn	626
aat Asn 205	cgc Arg	tgc Cys	aga Arg	cag Gln	gaa Glu 210	gat Asp	aca Thr	gag Glu	agg Arg	gta Val 215	cga Arg	gca Ala	gtg Val	aca Thr	aag Lys 220	674
cca Pro	aat Asn	aac Asn	gaa Glu	gaa Glu 225	agt Ser	tgt Cys	gaa Glu	aag Lys	cca Pro 230	cat His	aga Arg	gtg Val	atg Met	ccg Pro 235	aat Asn	722
ttt Phe	gct Ala	gaa Glu	gtt Val 240	tac Tyr	agc Ser	ttc Phe	att Ile	gga Gly 245	agt Ser	gtc Val	ttc Phe	gat Asp	ccc Pro 250	aac Asn	aca Thr	770
tca Ser	ggc Gly	cac His 255	ctc Leu	cag Gln	aga Arg	tta Leu	aag Lys 260	cag Gln	atg Met	gat Asp	cca Pro	ata Ile 265	aat Asn	atg Met	gaa Glu	818
acg Thr	gtt Val 270	ctt Leu	tta Leu	ctg Leu	atg Met	caa Gln 275	aac Asn	ctg Leu	tct Ser	gta Val	aat Asn 280	ctg Leu	aca Thr	agt Ser	ccc Pro	866
gag Glu 285	ttt Phe	gca Ala	gag Glu	caa Gln	agg Arg 290	agg Arg	ttg Leu	ata Ile	tca Ser	tca Ser 295	tac Tyr	agc Ser	gct Ala	aaa Lys	gct Ala 300	914
	aaa Lys	tag	agat	agaa	ata a	aaca	aataa	at gt	tacct	tate	g tga	agato	caag			963
aga	caato	cat o	caaç	ggtci	g ta	atgca	attgo	t t t q	ggatt	tag	gcci	cgt	gtt (	ctcad	ctacag	1023
gage	cagaa	acc a	aatc	gcaaa	ag ac	ctct	agat	ggd	ctact	gag	ttg	ggti	tt 1	tatg	ctctg	1083
taag	gtcg	egg t	ggaç	gcaca	ac gt	gtt	gtc	tgt	cttg	gtgt	atg	gtg	tat a	agata	aataca	1143
agg	ttt	gca ç	gagta	aaggt	c a	agti	agct	gca	aagto	gagt	ttg	gatca	aat d	cttaa	agatta	1203
aaa	cct	gag a	agtga	agtgi	c ca	aaaga	agacı	gtg	gtaat	att	ggti	tgg	egg (	cago	cagaag	1263
															cacaag	1323
ttt	agtti	ttg a	aataa	attgi	a ta	actca	aaaa	a ata	atgaa	atgt	ataa	aagaa	ata a	atcad	cttgag	1383
tcg	cctta	a														1391

#### mbi19 Sequence Listing.ST25

<210> 12 <211> 302 <212> PRT <213> Arabidopsis thaliana <400> 12 Met Val Ser Val Asn Pro Arg Pro Lys Gly Phe Pro Val Phe Asp Ser Ser Asn Met Ser Leu Pro Ser Ser Asp Gly Phe Gly Ser Ile Pro Ala 20 25 30Thr Gly Arg Thr Ser Thr Val Ser Phe Ser Glu Asp Pro Thr Thr Lys Ile Arg Lys Pro Tyr Thr Ile Lys Lys Ser Arg Glu Asn Trp Thr Asp 50 55 Gln Glu His Asp Lys Phe Leu Glu Ala Leu His Leu Phe Asp Arg Asp 65 70 75 80 Trp Lys Lys Ile Glu Ala Phe Val Gly Ser Lys Thr Val Val Gln Ile  $85 \hspace{1cm} 90 \hspace{1cm} 95$ Arg Ser His Ala Gln Lys Tyr Phe Leu Lys Val Gln Lys Ser Gly Ala 100 105 110 Asn Glu His Leu Pro Leu Pro Arg Pro Lys Arg Lys Ala Ser His Pro 115 120 125 Tyr Pro Ile Lys Ala Pro Lys Asn Val Ala Tyr Thr Ser Leu Pro Ser 130 135 140 Ser Ser Thr Leu Pro Leu Leu Glu Pro Gly Tyr Leu Tyr Ser Ser Asp 145 150 155 160 Ser Lys Ser Leu Met Gly Asn Gln Ala Val Cys Ala Ser Thr Ser Ser 165 170 175 Ser Trp Asn His Glu Ser Thr Asn Leu Pro Lys Pro Val Ile Glu Glu Glu Pro Gly Val Ser Ala Thr Ala Pro Leu Pro Asn Asn Arg Cys Arg 195 200 205 Gln Glu Asp Thr Glu Arg Val Arg Ala Val Thr Lys Pro Asn Asn Glu 210 215 220 Glu Ser Cys Glu Lys Pro His Arg Val Met Pro Asn Phe Ala Glu Val 225 230 240 Tyr Ser Phe Ile Gly Ser Val Phe Asp Pro Asn Thr Ser Gly His Leu 245 250 255 Gln Arg Leu Lys Gln Met Asp Pro Ile Asn Met Glu Thr Val Leu Leu

#### mbil9 Sequence Listing.ST25

Leu Met Gln Asn Leu Ser Val Asn Leu Thr Ser Pro Glu Phe Ala Glu 275 280 285

Gln Arg Arg Leu Ile Ser Ser Tyr Ser Ala Lys Ala Leu Lys 290 295 300

<210> 13 <211> 1121 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (48)..(902) <223> G263 <400> 13 tttttagttt tatttttctg tggtaaaata aaaaaagttc gccggag atg acg gct Met Thr Ala 56 gtg acg gcg gcg caa aga tca gtt ccg gcg ccg ttt tta agc aaa acg Val Thr Ala Ala Gln Arg Ser Val Pro Ala Pro Phe Leu Ser Lys Thr 5 10 15 104 tat cag cta gtt gat gat cat agc aca gac gac gtc gtt tca tgg aac Tyr Gln Leu Val Asp Asp His Ser Thr Asp Asp Val Val Ser Trp Asn 20 25 30 35152 gaa gaa gga aca gct ttt gtc gtg tgg aaa aca gca gag ttt gct aaa Glu Glu Gly Thr Ala Phe Val Val Trp Lys Thr Ala Glu Phe Ala Lys 200 gat ctt ctt cct caa tac ttc aag cat aat aat ttc tca agc ttc att Asp Leu Leu Pro Gln Tyr Phe Lys His Asn Asn Phe Ser Ser Phe Ile 248 cgt cag ctc aac act tac gga ttt cgt aaa act gta ccg gat aaa tgg Arg Gln Leu Asn Thr Tyr Gly Phe Arg Lys Thr Val Pro Asp Lys Trp 70 75 80 296 gaa ttt gca aac gat tat ttc cgg aga ggc ggg gag gat ctg ttg acg Glu Phe Ala Asn Asp Tyr Phe Arg Arg Gly Gly Glu Asp Leu Leu Thr 344 gac ata cga cgg cgt aaa tcg gtg att gct tca acg gcg ggg aaa tgt Asp Ile Arg Arg Lys Ser Val Ile Ala Ser Thr Ala Gly Lys Cys 392 gtt gtt gtt ggt tcg cct tct gag tct aat tct ggt ggt ggt gat gat Val Val Val Gly Ser Pro Ser Glu Ser Asn Ser Gly Gly Gly Asp Asp 440 cac ggt toa ago too acg toa toa coc ggt tog tog aag aat cot ggt His Gly Ser Ser Ser Thr Ser Ser Pro Gly Ser Ser Lys Asn Pro Gly 488 tcg gtg gag aac atg gtt gct gat tta tca gga gag aac gag aag ctt Ser Val Glu Asn Met Val Ala Asp Leu Ser Gly Glu Asn Glu Lys Leu 536 584 aaa cgt gaa aac aat aac ttg agc tcg gag ctc gcg gcg gcg aag aag Lys Arg Glu Asn Asn Asn Leu Ser Ser Glu Leu Ala Ala Ala Lys Lys 170 cag cgc gat gag cta gtg acg ttc ttg acg ggt cat ctg aaa gta aga Gln Arg Asp Glu Leu Val Thr Phe Leu Thr Gly His Leu Lys Val Arg 180 185 190 195 632 ccg gaa caa atc gat aaa atg atc aaa gga ggg aaa ttt aaa ccg gtg Page 15

PCT/US00/31414 WO 01/35725 mbi19 Sequence Listing.ST25
Pro Glu Gln Ile Asp Lys Met Ile Lys Gly Gly Lys Phe Lys Pro Val
200 205 210 gag tct gac gaa gag agt gag tgc gaa ggt tgc gac ggc ggc gga gga Glu Ser Asp Glu Glu Ser Glu Cys Glu Gly Cys Asp Gly Gly Gly 215 220 225 728 gca gag gag ggg gta ggt gaa gga ttg aaa ttg ttt ggg gtg tgg ttg Ala Glu Glu Gly Val Gly Glu Gly Leu Lys Leu Phe Gly Val Trp Leu 230 235 240 776 824 aaa gga gag aga aaa aag agg gac cgg gat gaa aag aat tat gtg gtg Lys Gly Glu Arg Lys Lys Arg Asp Arg Asp Glu Lys Asn Tyr Val Val 245 250 255 agt ggg tcc cgt atg acg gaa ata aag aac gtg gac ttt cac gcg ccg Ser Gly Ser Arg Met Thr Glu Ile Lys Asn Val Asp Phe His Ala Pro 872 ttg tgg aaa agc agc aaa gtc tgc aac taa aaaaagagta gaagactgtt Leu Trp Lys Ser Ser Lys Val Cys Asn 922 982 caaaccagcg tgtgacacgt catcgacgac gacgaaaaaa atgatttaaa aaactatttt 1042 tttccgtaag gaagaaaagt tatttttatg ttttaaaaaag gtgaagaagg tccagaagga 1102 tcaacgcaaa tatataaatg gattttcatg tattatataa tttaattagt gtattaagaa 1121 aataaaacaa aaaaaaaaa <210> 14 <211> 284 <212> PRT <213> Arabidopsis thaliana <400> 14 Met Thr Ala Val Thr Ala Ala Gln Arg Ser Val Pro Ala Pro Phe Leu 1 5 10 15 Ser Lys Thr Tyr Gln Leu Val Asp Asp His Ser Thr Asp Asp Val Val Ser Trp Asn Glu Glu Gly Thr Ala Phe Val Val Trp Lys Thr Ala Glu 35 40 45Phe Ala Lys Asp Leu Leu Pro Gln Tyr Phe Lys His Asn Asn Phe Ser 50 60 Ser Phe Ile Arg Gln Leu Asn Thr Tyr Gly Phe Arg Lys Thr Val Pro 65 75 80 Asp Lys Trp Glu Phe Ala Asn Asp Tyr Phe Arg Arg Gly Glu Asp

Gly Asp Asp His Gly Ser Ser Ser Thr Ser Ser Pro Gly Ser Ser Lys 130 135 140

Leu Leu Thr Asp Ile Arg Arg Arg Lys Ser Val Ile Ala Ser Thr Ala 100 105 110

Gly Lys Cys Val Val Val Gly Ser Pro Ser Glu Ser Asn Ser Gly Gly

mbi19 Sequence Listing.ST25
Asn Pro Gly Ser Val Glu Asn Met Val Ala Asp Leu Ser Gly Glu Asn Glu Lys Leu Lys Arg Glu Asn Asn Leu Ser Ser Glu Leu Ala Ala Ala Lys Lys Gln Arg Asp Glu Leu Val Thr Phe Leu Thr Gly His Leu 180 185 190 Lys Val Arg Pro Glu Gln Ile Asp Lys Met Ile Lys Gly Gly Lys Phe 195 200 205 Lys Pro Val Glu Ser Asp Glu Glu Ser Glu Cys Glu Gly Cys Asp Gly 210 215 220 Gly Gly Gly Ala Glu Glu Gly Val Gly Glu Gly Leu Lys Leu Phe Gly 225 230 240 Val Trp Leu Lys Gly Glu Arg Lys Lys Arg Asp Arg Asp Glu Lys Asn 245 250 255 Tyr Val Val Ser Gly Ser Arg Met Thr Glu Ile Lys Asn Val Asp Phe His Ala Pro Leu Trp Lys Ser Ser Lys Val Cys Asn <210> 1951 <211> <212> DNA <213> Arabidopsis thaliana <220> <221> CDS (196)..(1794) <222> G308 <400> 15 agtaatttag ttttttttt tttttttac aatttatttt gttattagaa gtggtagtgg 60 120 agtgaaaaaa caaatcctaa gcagtcctaa ccgatccccg aagctaaaga ttcttcacct 180 tcccaaataa agcaaaacct agatccgaca ttgaaggaaa aaccttttag atccatctct gaaaaaaacc caacc atg aag aga gat cat cat cat cat cat caa gat aag
Met Lys Arg Asp His His His His Gln Asp Lys 231 aag act atg atg atg aat gaa gac gac ggt aac ggc atg gat gag Lys Thr Met Met Met Asn Glu Glu Asp Asp Gly Asn Gly Met Asp Glu 20 ctt cta gct gtt ctt ggt tac aag gtt agg tca tcg gaa atg gct gat Leu Leu Ala Val Leu Gly Tyr Lys Val Arg Ser Ser Glu Met Ala Asp 30 35 40 327 gtt gct cag aaa ctc gag cag ctt gaa gtt atg atg tct aat gtt caa Val Ala Gln Lys Leu Glu Gln Leu Glu Val Met Met Ser Asn Val Gln 50 gaa gac gat ctt tct caa ctc gct act gag act gtt cac tat aat ccg Glu Asp Asp Leu Ser Gln Leu Ala Thr Glu Thr Val His Tyr Asn Pro 423 gcg gag ctt tac acg tgg ctt gat tct atg ctc acc gac ctt aat cct Page 17

Ala	Glu	Leu	Tyr 80	Thr	Trp	Leu					List Thr				Pro	
		tct Ser 95														519
		aat Asn														567
ggc Gly 125	gga Gly	gga Gly	gat Asp	acg Thr	tat Tyr 130	act Thr	aca Thr	aac Asn	aag Lys	cgg Arg 135	ttg Leu	aaa Lys	tgc Cys	tca Ser	aac Asn 140	615
ggc Gly	gtc Val	gtg Val	gaa Glu	acc Thr 145	acc Thr	aca Thr	gcg Ala	acg Thr	gct Ala 150	gag Glu	tca Ser	act Thr	cgg Arg	cat His 155	gtt Val	663
		gtt Val														711
ttg Leu	gct Ala	tgc Cys 175	gct Ala	gaa Glu	gct Ala	gtt Val	cag Gln 180	aag Lys	gag Glu	aat Asn	ctg Leu	act Thr 185	gtg Val	gcg Ala	gaa Glu	759
gct Ala	ctg Leu 190	gtg Val	aag Lys	caa Gln	atc Ile	gga Gly 195	ttc Phe	tta Leu	gct Ala	gtt Val	tct Ser 200	caa Gln	atc Ile	gga Gly	gct Ala	807
		caa Gln														855
tac Tyr	cgt Arg	ctc Leu	tct Ser	ccg Pro 225	tcg Ser	cag Gln	agt Ser	cca Pro	atc Ile 230	gac Asp	cac His	tct Ser	ctc Leu	tcc Ser 235	gat Asp	903
		cag Gln													gct Ala	951
		acg Thr 255														999
aga Arg	gtt Val 270	cat His	gtc Val	att Ile	gat Asp	ttc Phe 275	tct Ser	atg Met	agt Ser	caa Gln	ggt Gly 280	ctt Leu	caa Gln	tgg Trp	ccg Pro	1047
		atg Met														1095
cgg Arg	tta Leu	acc Thr	gga Gly	att Ile 305	ggt Gly	cca Pro	ccg Pro	gca Ala	ccg Pro 310	gat Asp	aat Asn	ttc Phe	gat Asp	tat Tyr 315	ctt Leu	1143
cat His	gaa Glu	gtt Val	999 Gly 320	tgt Cys	aag Lys	ctg Leu	gct Ala	cat His 325	tta Leu	gct Ala	gag Glu	gcg Ala	att Ile 330	cac His	gtt Val	1191
gag Glu	ttt Phe	gag Glu 335	tac Tyr	aga Arg	gga Gly	ttt Phe	gtg Val 340	gct Ala	aac Asn	act Thr	tta Leu	gct Ala 345	gat Asp	ctt Leu	gat Asp	1239
gct Ala	tcg Ser 350	atg Met	ctt Leu	gag Glu	ctt Leu	aga Arg 355	cca Pro	agt Ser	gag Glu	att Ile	gaa Glu 360	tct Ser	gtt Val	gcg Ala	gtt Val	1287
aac Asn 365	tct Ser	gtt Val	ttc Phe	gag Glu	ctt Leu 370	cac His	aag Lys	ctc Leu	ttg Leu	gga Gly 375	cga Arg	cct Pro	ggt Gly	gcg Ala	atc Ile 380	1335

gat Asp	aag Lys	gtt Val	ctt Leu	ggt Gly 385	gtg Val	gtg Val	aat	caq	eque att Ile 390	aaa	CCq	gag	att	ttc Phe 395	act Thr	1383
gtg Val	gtt Val	gag Glu	cag Gln 400	gaa Glu	tcg Ser	aac Asn	cat His	aat Asn 405	agt Ser	ccg Pro	att Ile	ttc Phe	tta Leu 410	gat Asp	cgg Arg	1431
ttt Phe	act Thr	gag Glu 415	tcg Ser	ttg Leu	cat His	tat Tyr	tac Tyr 420	tcg Ser	acg Thr	ttg Leu	ttt Phe	gac Asp 425	tcg Ser	ttg Leu	gaa Glu	1479
ggt Gly	gta Val 430	ccg Pro	agt Ser	ggt Gly	caa Gln	gac Asp 435	aag Lys	gtc Val	atg Met	tcg Ser	gag Glu 440	gtt Val	tac Tyr	ttg Leu	ggt Gly	1527
aaa Lys 445	cag Gln	atc Ile	tgc Cys	aac Asn	gtt Val 450	gtg Val	gct Ala	tgt Cys	gat Asp	gga Gly 455	cct Pro	gac Asp	cga Arg	gtt Val	gag Glu 460	1575
cgt Arg	cat His	gaa Glu	acg Thr	ttg Leu 465	agt Ser	cag Gln	tgg Trp	agg Arg	aac Asn 470	cgg Arg	ttc Phe	999 Gly	tct Ser	gct Ala 475	ggg Gly	1623
ttt Phe	gcg Ala	gct Ala	gca Ala 480	cat His	att Ile	ggt Gly	tcg Ser	aat Asn 485	gcg Ala	ttt Phe	aag Lys	caa Gln	gcg Ala 490	agt Ser	atg Met	1671
ctt Leu	ttg Leu	gct Ala 495	ctg Leu	ttc Phe	aac Asn	ggc Gly	ggt Gly 500	gag Glu	ggt Gly	tat Tyr	cgg Arg	gtg Val 505	gag Glu	gag Glu	agt Ser	1719
gac Asp	ggc Gly 510	tgt Cys	ctc Leu	atg Met	ttg Leu	ggt Gly 515	tgg Trp	cac His	aca Thr	cga Arg	ccg Pro 520	ctc Leu	ata Ile	gcc Ala	acc Thr	1767
	gct Ala							tag	atg	gtggd	etc a	atga	aattg	ga		1814
tct	gttga	aac o	ggtt	atga	at ga	ataga	attt	c cga	accga	agc	caaa	actaa	aat o	ccta	etgttt	1874
ttc	ctt	gt o	cactt	gtta	aa ga	atctt	atc	tto	catta	tat	tagg	gtaat	tg a	aaaa	atttta	1934
atc	tege	cta a	aatta	act												1951
<21 <21 <21 <21	1> ! 2> !	16 532 PRT Arabi	idops	sis t	hali	iana										
<40	0> :	L6														
Met 1	Lys	Arg	Asp	His 5	His	His	His	His	Gln 10	Asp	Lys	Lys	Thr	Met 15	Met	
Met	Asn	Glu	Glu 20	Asp	Asp	Gly	Asn	Gly 25	Met	Asp	Glu	Leu	Leu 30	Ala	Val	

Leu Gly Tyr Lys Val Arg Ser Ser Glu Met Ala Asp Val Ala Gln Lys 35

Leu Glu Gln Leu Glu Val Met Met Ser Asn Val Gln Glu Asp Asp Leu 50 60

Ser Gln Leu Ala Thr Glu Thr Val His Tyr Asn Pro Ala Glu Leu Tyr 65 70 80

Thr Trp Leu Asp Ser Met Leu Thr Asp Leu Asn Pro Pro Ser Ser Asn Page 19

mbil9 Sequence Listing.ST25 90 95

Ala Glu Tyr Asp Leu Lys Ala Ile Pro Gly Asp Ala Ile Leu Asn Gln 100 105 110

Phe Ala Ile Asp Ser Ala Ser Ser Ser Asn Gln Gly Gly Gly Asp 115 120 125

Thr Tyr Thr Thr Asn Lys Arg Leu Lys Cys Ser Asn Gly Val Val Glu 130 135 140

Thr Thr Thr Ala Thr Ala Glu Ser Thr Arg His Val Val Leu Val Asp 145 150 160

Ser Gln Glu Asn Gly Val Arg Leu Val His Ala Leu Leu Ala Cys Ala 165 170 175

Glu Ala Val Gln Lys Glu Asn Leu Thr Val Ala Glu Ala Leu Val Lys 180 185 190

Gln Ile Gly Phe Leu Ala Val Ser Gln Ile Gly Ala Met Arg Gln Val 195 200 205

Ala Thr Tyr Phe Ala Glu Ala Leu Ala Arg Arg Ile Tyr Arg Leu Ser 210 215 220

Pro Ser Gln Ser Pro Ile Asp His Ser Leu Ser Asp Thr Leu Gln Met 225 230 240

His Phe Tyr Glu Thr Cys Pro Tyr Leu Lys Phe Ala His Phe Thr Ala 245 250 255

Asn Gln Ala Ile Leu Glu Ala Phe Gln Gly Lys Lys Arg Val His Val 260 265 270

Ile Asp Phe Ser Met Ser Gln Gly Leu Gln Trp Pro Ala Leu Met Gln 275 280 285

Ala Leu Ala Leu Arg Pro Gly Gly Pro Pro Val Phe Arg Leu Thr Gly 290 295 300

Ile Gly Pro Pro Ala Pro Asp Asn Phe Asp Tyr Leu His Glu Val Gly 305 310 320

Cys Lys Leu Ala His Leu Ala Glu Ala Ile His Val Glu Phe Glu Tyr 325 330 335

Arg Gly Phe Val Ala Asn Thr Leu Ala Asp Leu Asp Ala Ser Met Leu 340 345 350

Glu Leu Arg Pro Ser Glu Ile Glu Ser Val Ala Val Asn Ser Val Phe

Glu Leu His Lys Leu Leu Gly Arg Pro Gly Ala Ile Asp Lys Val Leu 370 380

mbil9 Sequence Listing.ST25
Gly Val Val Asn Gln Ile Lys Pro Glu Ile Phe Thr Val Val Glu 400 395 Glu Ser Asn His Asn Ser Pro Ile Phe Leu Asp Arg Phe Thr Glu Ser Leu His Tyr Tyr Ser Thr Leu Phe Asp Ser Leu Glu Gly Val Pro Ser 420 425 430Gly Gln Asp Lys Val Met Ser Glu Val Tyr Leu Gly Lys Gln Ile Cys Asn Val Val Ala Cys Asp Gly Pro Asp Arg Val Glu Arg His Glu Thr Leu Ser Gln Trp Arg Asn Arg Phe Gly Ser Ala Gly Phe Ala Ala Ala His Ile Gly Ser Asn Ala Phe Lys Gln Ala Ser Met Leu Leu Ala Leu Phe Asn Gly Gly Glu Gly Tyr Arg Val Glu Glu Ser Asp Gly Cys Leu Met Leu Gly Trp His Thr Arg Pro Leu Ile Ala Thr Ser Ala Trp Lys Leu Ser Thr Asn <210> 17 <211> 768 <212> DNA <213> Arabidopsis thaliana <221> CDS <222> (1)..(768) <223> G536 atg tcg aca agg gaa gag aat gtt tac atg gcg aaa tta gcc gaa caa Met Ser Thr Arg Glu Glu Asn Val Tyr Met Ala Lys Leu Ala Glu Gln 1 5 15 48 gct gaa cgt tac gaa gaa atg gtt gaa ttc atg gag aaa gtt gcg aaa Ala Glu Arg Tyr Glu Glu Met Val Glu Phe Met Glu Lys Val Ala Lys 96 act gtt gat gtt gag gaa ctt tca gtt gaa gag agg aat ctt ctc tct Thr Val Asp Val Glu Glu Leu Ser Val Glu Glu Arg Asn Leu Leu Ser 144 gtt gct tac aag aac gtg att gga gcg aga aga gct tcg tgg aga atc Val Ala Tyr Lys Asn Val Ile Gly Ala Arg Arg Ala Ser Trp Arg Ile 50 55 192 att tct tcg att gag cag aaa gaa gag agc aaa ggg aac gaa gat cat Ile Ser Ser Ile Glu Gln Lys Glu Glu Ser Lys Gly Asn Glu Asp His 65 70 75 80 240 gtt gct att atc aag gat tac aga gga gag att gaa tcc gag ctt agc Val Ala Ile Ile Lys Asp Tyr Arg Gly Glu Ile Glu Ser Glu Leu Ser 85 90 95 288

PCT/US00/31414 WO 01/35725

### mbil9 Sequence Listing.ST25

									•			_				
aaa Lys	atc Ile	tgt Cys	gat Asp 100	999 Gly	att Ile	ttg Leu	aat Asn	gtt Val 105	ctt Leu	gaa Glu	gct Ala	cat His	ctt Leu 110	att Ile	cct Pro	336
tct Ser	gct Ala	tca Ser 115	cca Pro	gct Ala	gaa Glu	tct Ser	aaa Lys 120	gtg Val	ttt Phe	tat Tyr	ctt Leu	aag Lys 125	atg Met	aag Lys	ggt Gly	384
gat Asp	tat Tyr 130	cat His	agg Arg	tat Tyr	ctt Leu	gct Ala 135	gag Glu	ttt Phe	aag Lys	gct Ala	ggt Gly 140	gct Ala	gaa Glu	agg Arg	aaa Lys	432
gaa Glu 145	gct Ala	gct Ala	gaa Glu	agc Ser	act Thr 150	ttg Leu	gtt Val	gct Ala	tac Tyr	aag Lys 155	tct Ser	gct Ala	tcc Ser	gac Asp	att Ile 160	480
gcc Ala	act Thr	gct Ala	gag Glu	tta Leu 165	gct Ala	cct Pro	act Thr	cac His	ccg Pro 170	ata Ile	agg Arg	ctt Leu	ggt Gly	ctt Leu 175	gca Ala	528
ctc Leu	aac Asn	ttc Phe	tct Ser 180	gtg Val	ttt Phe	tac Tyr	tat Tyr	gaa Glu 185	atc	ctc Leu	aac Asn	tcg Ser	cct Pro 190	gat Asp	cgt Arg	576
gct Ala	tgc Cys	agc Ser 195	ctc Leu	gca Ala	aag Lys	cag Gln	gcg Ala 200	ttt Phe	gat Asp	gat Asp	gça Ala	atc Ile 205	gct Ala	gag Glu	tta Leu	624
gat Asp	aca Thr 210	ttg Leu	ggt Gly	gag Glu	gaa Glu	tca Ser 215	tac Tyr	aag Lys	gac Asp	agt Ser	aca Thr 220	ctg Leu	att Ile	atg Met	cag Gln	672
ctt Leu 225	ctt Leu	aga Arg	gac Asp	aat Asn	ctc Leu 230	act Thr	ctc Leu	tgg Trp	act Thr	tca Ser 235	gat Asp	atg Met	act Thr	gac Asp	gaa Glu 240	720
gca Ala	gga Gly	gat Asp	gag Glu	att Ile 245	aag Lys	gag Glu	gca Ala	tca Ser	aag Lys 250	ccc Pro	gat Asp	ggt Gly	gcc Ala	gag Glu 255	taa	768

<210> 18 <211> 255 <212> PRT <213> Arabidopsis thaliana

<400> 18

Ala Glu Arg Tyr Glu Glu Met Val Glu Phe Met Glu Lys Val Ala Lys 20 30

Thr Val Asp Val Glu Glu Leu Ser Val Glu Glu Arg Asn Leu Leu Ser 35 40 45

Val Ala Tyr Lys Asn Val Ile Gly Ala Arg Arg Ala Ser Trp Arg Ile 50 60

Ile Ser Ser Ile Glu Gln Lys Glu Glu Ser Lys Gly Asn Glu Asp His 65 70 75 80

Val Ala Ile Ile Lys Asp Tyr Arg Gly Glu Ile Glu Ser Glu Leu Ser 85 90 95

Lys Ile Cys Asp Gly Ile Leu Asn Val Leu Glu Ala His Leu Ile Pro  $100 \hspace{1cm} 105 \hspace{1cm} 110$ 

### mbil9 Sequence Listing.ST25

Ser Ala Ser Pro Ala Glu Ser Lys Val Phe Tyr Leu Lys Met Lys Gly Asp Tyr His Arg Tyr Leu Ala Glu Phe Lys Ala Gly Ala Glu Arg Lys 130 135 140 Glu Ala Ala Glu Ser Thr Leu Val Ala Tyr Lys Ser Ala Ser Asp Ile Ala Thr Ala Glu Leu Ala Pro Thr His Pro Ile Arg Leu Gly Leu Ala Leu Asn Phe Ser Val Phe Tyr Tyr Glu Ile Leu Asn Ser Pro Asp Arg Ala Cys Ser Leu Ala Lys Gln Ala Phe Asp Asp Ala Ile Ala Glu Leu Asp Thr Leu Gly Glu Glu Ser Tyr Lys Asp Ser Thr Leu Ile Met Gln Leu Leu Arg Asp Asn Leu Thr Leu Trp Thr Ser Asp Met Thr Asp Glu Ala Gly Asp Glu Ile Lys Glu Ala Ser Lys Pro Asp Gly Ala Glu <210> 19 <211> 2526 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (338)..(2275) <400> 19 cagttatett etteettett etetetgttt tttaaattta tttttagaga attttttttg 60 ttttgcttcc gatttgatta tttccgggaa cgatgacttc tccggggagt tcccggtgag 120 atgataagtc agattgcata cttgtctcct ccatggctac tctcaagggt tttggctgcg 180 gtggattcgt ttggtttctc tagaatctaa agaggttatc acaacggctt tgcaatttga 240 aaactttcat gtttggggag atcaaagatg gtttcttttt tatactttac ttgttagaga 300 355 ggatttgaag cagcgaatag ctgcaaccgg teetgtt atg gat act aat aca tet Met Asp Thr Asn Thr Ser gga gaa gaa tta tta gct aag gca aga aag cca tat aca ata aca aag Cly Glu Glu Leu Ala Lys Ala Arg Lys Pro Tyr Thr Ile Thr Lys 403 cag cga gag cga tgg act gag gat gag cat gag agg ttt cta gaa gcc Gln Arg Glu Arg Trp Thr Glu Asp Glu His Glu Arg Phe Leu Glu Ala 451 ttg agg ctt tat gga aga gct tgg caa cga att gaa gaa cat att ggg Leu Arg Leu Tyr Gly Arg Ala Trp Gln Arg Ile Glu Glu His Ile Gly 40 45 50499

# mbi19 Sequence Listing.ST25

aca Thr 55	aag Lys	act Thr	gct Ala	gtt Val	cag Gln 60	atc Ile	aga Arg	agt Ser	cat His	gca Ala 65	caa Gln	aag Lys	ttc Phe	ttc Phe	aca Thr 70	547
aag Lys	ttg Leu	gag Glu	aaa Lys	gag Glu 75	gct Ala	gaa Glu	gtt Val	aaa Lys	ggc Gly 80	atc Ile	cct Pro	gtt Val	tgc Cys	caa Gln 85	gct Ala	595
ttg Leu	gac Asp	ata Ile	gaa Glu 90	att Ile	ccg Pro	cct Pro	cct Pro	cgt Arg 95	cct Pro	aaa Lys	cga Arg	aaa Lys	ccc Pro 100	aat Asn	act Thr	643
cct Pro	tat Tyr	cct Pro 105	cga Arg	aaa Lys	cct Pro	gly Gly	aac Asn 110	aac Asn	ggt Gly	aca Thr	tct Ser	tcc Ser 115	tct Ser	caa Gln	gta Val	691
tca Ser	tca Ser 120	gca Ala	aaa Lys	gat Asp	gca Ala	aaa Lys 125	ctt Leu	gtt Val	tca Ser	tcg Ser	gcc Ala 130	tct Ser	tct Ser	tca Ser	cag Gln	739
ttg Leu 135	aat Asn	cag Gln	gcg Ala	ttc Phe	ttg Leu 140	gat Asp	ttg Leu	gaa Glu	aaa Lys	atg Met 145	ccg Pro	ttc Phe	tct Ser	gag Glu	aaa Lys 150	787
aca Thr	tca Ser	act Thr	gga Gly	aaa Lys 155	gaa Glu	aat Asn	caa Gln	gat Asp	gag Glu 160	aat Asn	tgc Cys	tcg Ser	ggt Gly	gtt Val 165	tct Ser	835
act Thr	gtg Val	aac Asn	aag Lys 170	tat Tyr	ccc Pro	tta Leu	cca Pro	acg Thr 175	aaa Lys	cag Gln	gta Val	agt Ser	ggc Gly 180	gac Asp	att Ile	883
gaa Glu	aca Thr	agt Ser 185	aag Lys	acc Thr	tca Ser	act Thr	gtg Val 190	gac Asp	aac Asn	gcg Ala	gtt Val	caa Gln 195	gat Asp	gtt Val	ccc Pro	931
aag Lys	aag Lys 200	aac Asn	aaa Lys	gac Asp	aaa Lys	gat Asp 205	ggt Gly	aac Asn	gat Asp	ggt Gly	act Thr 210	act Thr	gtg Val	cac His	agc Ser	979
atg Met 215	caa Gln	aac Asn	tac Tyr	cct Pro	tgg Trp 220	cat His	ttc Phe	cac His	gca Ala	gat Asp 225	att Ile	gtg Val	aac Asn	999 999	aat Asn 230	1027
ata Ile	gca Ala	aaa Lys	tgc Cys	cct Pro 235	caa Gln	aat Asn	cat His	ccc Pro	tca Ser 240	ggt Gly	atg Met	gta Val	tct Ser	caa Gln 245	gac Asp	1075
ttc Phe	atg Met	ttt Phe	cat His 250	cct Pro	atg Met	aga Arg	gaa Glu	gaa Glu 255	act Thr	cac His	ggg Gly	cac His	gca Ala 260	aat Asn	ctt Leu	1123
caa Gln	gct Ala	aca Thr 265	aca Thr	gca Ala	tct Ser	gct Ala	act Thr 270	act Thr	aca Thr	gct Ala	tct Ser	cat His 275	caa Gln	gcg Ala	ttt Phe	1171
cca Pro	gct Ala 280	tgt Cys	cat His	tca Ser	cag Gln	gat Asp 285	gat Asp	tac Tyr	cgt Arg	tcg Ser	ttt Phe 290	ctc Leu	cag Gln	ata Ile	tca Ser	1219
tct Ser 295	act Thr	ttc Phe	tcc Ser	aat Asn	ctt Leu 300	att Ile	atg Met	tca Ser	act Thr	ctc Leu 305	cta Leu	cag Gln	aat Asn	cct Pro	gca Ala 310	1267
gct Ala	cat His	gct Ala	gca Ala	gct Ala 315	aca Thr	ttc Phe	gct Ala	gct Ala	tcg Ser 320	gtc Val	tgg Trp	cct Pro	tat Tyr	gcg Ala 325	agt Ser	1315
gtc Val	ggg Gly	aat Asn	tct Ser 330	ggt Gly	gat Asp	tca Ser	tca Ser	acc Thr 335	cca Pro	atg Met	agc Ser	tct Ser	tct Ser 340	cct Pro	cca Pro	1363
agt Ser	ata Ile	act Thr	gcc Ala	att Ile	gcc Ala	gct Ala	gct Ala	aca Thr	Val	gct Ala age	Ala	gca Ala	act Thr	gct Ala	tgg Trp	1411

		345					mbi 350	19 S	eque	nce	List	ing. 355	ST25			
	gct Ala 360															1459
tgt Cys 375	gtt Val	cca Pro	ttc Phe	tca Ser	act Thr 380	gtt Val	gca Ala	gtt Val	cca Pro	act Thr 385	cca Pro	gca Ala	atg Met	act Thr	gaa Glu 390	1507
atg Met	gat Asp	acc Thr	gtt Val	gaa Glu 395	aat Asn	act Thr	caa Gln	ccg Pro	ttt Phe 400	gag Glu	aaa Lys	caa Gln	aac Asn	aca Thr 405	gct Ala	1555
ctg Leu	caa Gln	gat Asp	caa Gln 410	acc Thr	ttg Leu	gct Ala	tcg Ser	aaa Lys 415	tct Ser	cca Pro	gct Ala	tca Ser	tca Ser 420	tct Ser	gat Asp	1603
	tca Ser															1651
aat Asn	gat Asp 440	gat Asp	aaa Lys	att Ile	gag Glu	gag Glu 445	gtt Val	gtt Val	gtt Val	act Thr	gcc Ala 450	gct Ala	gtg Val	cat His	gac Asp	1699
tca Ser 455	aac Asn	act Thr	gcc Ala	cag Gln	aag Lys 460	aaa Lys	aat Asn	ctt Leu	gtg Val	gac Asp 465	cgc Arg	tca Ser	tcg Ser	tgt Cys	ggc Gly 470	1747
tca Ser	aat Asn	aca Thr	cct Pro	tca Ser 475	gly ggg	agt Ser	gac Asp	gca Ala	gaa Glu 480	act Thr	gat Asp	gca Ala	tta Leu	gat Asp 485	aaa Lys	1795
	gag Glu								Glu							1843
	gtt Val															1891
	aac Asn 520															1939
ggt Gly 535	cgt Arg	ata Ile	gcg Ala	ttt Phe	cag Gln 540	gct Ala	ctc Leu	ttt Phe	gca Ala	aga Arg 545	gaa Glu	aga Arg	ttg Leu	cct Pro	caa Gln 550	1987
agc Ser	ttt Phe	tcg Ser	cct Pro	cct Pro 555	caa Gln	gtg Val	gca Ala	gag Glu	aat Asn 560	gtg Val	aat Asn	aga Arg	aaa Lys	caa Gln 565	agt Ser	2035
	acg Thr															2083
	gca Ala															2131
	ctt Leu 600															2179
	gaa Glu															2227
	aaa Lys														tga	2275
caga	actto	gga g	ggtaa	aaaa	aa aa	acat	ccad	att		atca age		ctt	aa a	tcta	agtgtt	2335

mbi19	Sequence	Listing.ST25
-------	----------	--------------

agtagtttgc ttctccaatc tttatgaaag agactttaa ttttccttcc gaacatttct 2395
ttggtcatgt caggttctgt accatattac cccatgtctt gtctcttgtc tctgtttgtg 2455
tatgctactt gtggtctata tgtcatctgc tactactgtt aattaaccat taagcaatgg 2515
atttgtcttt a 2526

<210> 20 <211> 645

<212> PRT

<213> Arabidopsis thaliana

<400> 20

Met Asp Thr Asn Thr Ser Gly Glu Glu Leu Leu Ala Lys Ala Arg Lys 1 5 10 15

Pro Tyr Thr Ile Thr Lys Gln Arg Glu Arg Trp Thr Glu Asp Glu His 20 25 30

Glu Arg Phe Leu Glu Ala Leu Arg Leu Tyr Gly Arg Ala Trp Gln Arg 40 45

Ile Glu Glu His Ile Gly Thr Lys Thr Ala Val Gln Ile Arg Ser His 50 60

Ala Gln Lys Phe Phe Thr Lys Leu Glu Lys Glu Ala Glu Val Lys Gly 65 70 75 80

Ile Pro Val Cys Gln Ala Leu Asp Ile Glu Ile Pro Pro Pro Arg Pro 85 90 95

Lys Arg Lys Pro Asn Thr Pro Tyr Pro Arg Lys Pro Gly Asn Asn Gly 100 105 110

Thr Ser Ser Ser Gln Val Ser Ser Ala Lys Asp Ala Lys Leu Val Ser

Ser Ala Ser Ser Ser Gln Leu Asn Gln Ala Phe Leu Asp Leu Glu Lys 130 135 140

Met Pro Phe Ser Glu Lys Thr Ser Thr Gly Lys Glu Asn Gln Asp Glu 145 150 160

Asn Cys Ser Gly Val Ser Thr Val Asn Lys Tyr Pro Leu Pro Thr Lys 165 170 175

Gln Val Ser Gly Asp Ile Glu Thr Ser Lys Thr Ser Thr Val Asp Asn 180 185 190

Ala Val Gln Asp Val Pro Lys Lys Asn Lys Asp Lys Asp Gly Asn Asp 195 200 205

Gly Thr Thr Val His Ser Met Gln Asn Tyr Pro Trp His Phe His Ala 210 215 220

Asp Ile Val Asn Gly Asn Ile Ala Lys Cys Pro Gln Asn His Pro Ser

### mbil9 Sequence Listing.ST25

Gly Met Val Ser Gln Asp Phe Met Phe His Pro Met Arg Glu Glu Thr
245 250 255

His Gly His Ala Asn Leu Gln Ala Thr Thr Ala Ser Ala Thr Thr Thr 260 265 270

Ala Ser His Gln Ala Phe Pro Ala Cys His Ser Gln Asp Asp Tyr Arg 275 280 285

Ser Phe Leu Gln Ile Ser Ser Thr Phe Ser Asn Leu Ile Met Ser Thr 290 295 300

Leu Leu Gln Asn Pro Ala Ala His Ala Ala Ala Thr Phe Ala Ala Ser 305 310 315 320

Val Trp Pro Tyr Ala Ser Val Gly Asn Ser Gly Asp Ser Ser Thr Pro 325 330 335

Met Ser Ser Pro Pro Ser Ile Thr Ala Ile Ala Ala Ala Thr Val

Ala Ala Ala Thr Ala Trp Trp Ala Ser His Gly Leu Leu Pro Val Cys 355 360 365

Ala Pro Ala Pro Ile Thr Cys Val Pro Phe Ser Thr Val Ala Val Pro 370 375 380

Thr Pro Ala Met Thr Glu Met Asp Thr Val Glu Asn Thr Gln Pro Phe 385 390 395 400

Glu Lys Gln Asn Thr Ala Leu Gln Asp Gln Thr Leu Ala Ser Lys Ser 405 410 415

Pro Ala Ser Ser Ser Asp Asp Ser Asp Glu Thr Gly Val Thr Lys Leu 420 425 430

Asn Ala Asp Ser Lys Thr Asn Asp Asp Lys Ile Glu Glu Val Val 445 440 445

Thr Ala Ala Val His Asp Ser Asn Thr Ala Gln Lys Lys Asn Leu Val 450 455 460

Asp Arg Ser Ser Cys Gly Ser Asn Thr Pro Ser Gly Ser Asp Ala Glu 465 470 475 480

Thr Asp Ala Leu Asp Lys Met Glu Lys Asp Lys Glu Asp Val Lys Glu
485 490 495

Thr Asp Glu Asn Gln Pro Asp Val Ile Glu Leu Asn Asn Arg Lys Ile 500 505 510

Lys Met Arg Asp Asn Asn Ser Asn Asn Asn Ala Thr Thr Asp Ser Trp 515 525

Lys Glu Val Ser Glu Glu Gly Arg Ile Ala Phe Gln Ala Leu Phe Ala Page 27

mbil9 Sequence Listing.ST25

530 Arg Glu Arg Leu Pro Gln Ser Phe Ser Pro Pro Gln Val Ala Glu Asn Val Asn Arg Lys Gln Ser Asp Thr Ser Met Pro Leu Ala Pro Asn Phe Lys Ser Gln Asp Ser Cys Ala Ala Asp Gln Glu Gly Val Val Met Ile 580 585 590 Gly Val Gly Thr Cys Lys Ser Leu Lys Thr Arg Gln Thr Gly Phe Lys Pro Tyr Lys Arg Cys Ser Met Glu Val Lys Glu Ser Gln Val Gly Asn Ile Asn Asn Gln Ser Asp Glu Lys Val Cys Lys Arg Leu Arg Leu Glu Gly Glu Ala Ser Thr <210> 21 <211> 1281 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (64)..(1098) G867 <400> 21 cacaacacaa acacatttct gttttctcca ttgtttcaaa ccataaaaaa aaacacagat taa atg gaa tcg agt agc gtt gat gag agt act aca agt aca ggt tcc Met Glu Ser Ser Ser Val Asp Glu Ser Thr Thr Ser Thr Gly Ser 108 atc tgt gaa acc ccg gcg ata act ccg gcg aaa aag tcg tcg gta ggt lle Cys Glu Thr Pro Ala lle Thr Pro Ala Lys Lys Ser Ser Val Gly 20 25 30 156 204 aac tta tac agg atg gga agc gga tca agc gtt gtg tta gat tca gag Asn Leu Tyr Arg Met Gly Ser Gly Ser Ser Val Val Leu Asp Ser Glu aac ggc gta gaa gct gaa tct agg aag ctt ccg tcg tca aaa tac aaa Asn Gly Val Glu Ala Glu Ser Arg Lys Leu Pro Ser Ser Lys Tyr Lys 50 55 60252 ggt gtg gtg cca caa cca aac gga aga tgg gga gct cag att tac gag Gly Val Val Pro Gln Pro Asn Gly Arg Trp Gly Ala Gln Ile Tyr Glu 300 aaa cac cag cgc gtg tgg ctc ggg aca ttc aac gaa gaa gac gaa gcc Lys His Gln Arg Val Trp Leu Gly Thr Phe Asn Glu Glu Asp Glu Ala 348 gct cgt gcc tac gac gtc gcg gtt cac agg ttc cgt cgc cgt gac gcc Ala Arg Ala Tyr Asp Val Ala Val His Arg Phe Arg Arg Arg Asp Ala 396 gtc aca aat ttc aaa gac gtg aag atg gac gaa gac gag gtc gat ttc Val Thr Asn Phe Lys Asp Val Lys Met Asp Glu Asp Glu Val Asp Phe 444

	115			mbi	19 S 120	eque	nce	List	ing.	ST25 125			
ttg aat Leu Asn	tct cat Ser His 130	tcg a	aaa tc Lys Se	gag Glu 135	atc Ile	gtt Val	gat Asp	atg Met	ttg Leu 140	agg Arg	aaa Lys	cat His	492
act tat Thr Tyr 145	aac gaa Asn Glu	gag (	tta ga Leu Gl 15	ı Gln	agt Ser	aaa Lys	cgg Arg	cgt Arg 155	cgt Arg	aat Asn	ggt Gly	aac Asn	540
gga aac Gly Asn 160	atg act Met Thr	Arg '	acg tt Thr Le 165	g tta ı Leu	acg Thr	tcg Ser	999 Gly 170	ttg Leu	agt Ser	aat Asn	gat Asp	ggt Gly 175	588
gtt tct Val Ser	acg acg Thr Thr	999 Gly 180	ttt ag Phe Ar	a tcg g Ser	gcg Ala	gag Glu 185	gca Ala	ctg Leu	ttt Phe	gag Glu	aaa Lys 190	gcg Ala	636
gta acg Val Thr	cca ago Pro Ser 195	' qaA	gtt gg Val Gl	g aag y Lys	cta Leu 200	aac Asn	cgt Arg	ttg Leu	gtt Val	ata Ile 205	ccg Pro	aaa Lys	684
cat cac His His	gca gag Ala Glu 210	aaa d Lys l	cat tt His Ph	ccg Pro 215	tta Leu	ccg Pro	tca Ser	agt Ser	aac Asn 220	gtt Val	tcc Ser	gtg Val	732
aaa gga Lys Gly 225	gtg ttg Val Leu	ttg a	aac tt Asn Ph 23	e Glu	gac Asp	gtt Val	aac Asn	999 Gly 235	aaa Lys	gtg Val	tgg Trp	agg Arg	780
ttc cgt Phe Arg 240	tac tcg Tyr Ser	Tyr '	tgg aa Trp As 245	agt n Ser	agt Ser	cag Gln	agt Ser 250	tat Tyr	gtt Val	ttg Leu	act Thr	aaa Lys 255	828
ggt tgg Gly Trp	agc agg Ser Arg	ttc q Phe 1 260	gtt aa Val Ly	g gag s Glu	aag Lys	aat Asn 265	cta Leu	cgt Arg	gct Ala	ggt Gly	gac Asp 270	gtg Val	876
gtt agt Val Ser	ttc agt Phe Ser 275	Arg :	tct aa Ser As	ggt Gly	cag Gln 280	gat Asp	caa Gln	cag Gln	ttg Leu	tac Tyr 285	att Ile	999 Gly	924
tgg aag Trp Lys	tcg aga Ser Arg 290	tcc g	ggg tc Gly Se	a gat r Asp 295	tta Leu	gat Asp	gcg Ala	ggt Gly	cgg Arg 300	gtt Val	ttg Leu	aga Arg	972
ttg ttc Leu Phe 305	gga gtt Gly Val	aac a Asn :	att tc Ile Se 31	r Pro	gag Glu	agt Ser	tca Ser	aga Arg 315	aac Asn	gac Asp	gtc Val	gta Val	1020
gga aac Gly Asn 320	Lys Arg	gtg a	Asn As	o Thr	gag Glu	Met	Leu	tcg Ser	ttg Leu	gtg Val	tgt Cys	agc Ser 335	1068
aag aag Lys Lys	caa cgc Gln Arg	atc i Ile i 340	ttt ca Phe Hi	gcc Ala	tcg Ser	taa	caa	ctcti	ct (	tctti	tttt	tt	1118
tcttttgt	tg tttt	aataai	t tttt	aaaa	c tc	catt	ttcg	ttt	ctt	tat 1	ttgca	atcggt	1178
ttctttct	tc ttgt	ttacca	a aagg	tcat	g agi	tgti	ttt	gtt	gtati	tga 1	tgaad	ctgtaa	1238
attttatt	ta tagg	ataaa	t ttta	aaaaa	a aaa	aaaa	aaaa	aaa					1281
<211> 3 <212> P	2 44 RT rabidop	sis tl	halian	a									

<400> 22

mbi19 Sequence Listing.ST25

Cys Glu Thr Pro Ala Ile Thr Pro Ala Lys Lys Ser Ser Val Gly Asn 20 25 30

Leu Tyr Arg Met Gly Ser Gly Ser Ser Val Val Leu Asp Ser Glu Asn 35  $\phantom{\bigg|}40\phantom{\bigg|}40\phantom{\bigg|}45\phantom{\bigg|}$ 

Gly Val Glu Ala Glu Ser Arg Lys Leu Pro Ser Ser Lys Tyr Lys Gly
50 60

Val Val Pro Gln Pro Asn Gly Arg Trp Gly Ala Gln Ile Tyr Glu Lys 65 70 75 80

His Gln Arg Val Trp Leu Gly Thr Phe Asn Glu Glu Asp Glu Ala Ala 85 90 95

Arg Ala Tyr Asp Val Ala Val His Arg Phe Arg Arg Arg Asp Ala Val 100 105 110

Thr Asn Phe Lys Asp Val Lys Met Asp Glu Asp Glu Val Asp Phe Leu 115 120 125

Asn Ser His Ser Lys Ser Glu Ile Val Asp Met Leu Arg Lys His Thr 130 135 140

Tyr Asn Glu Glu Leu Glu Gln Ser Lys Arg Arg Arg Asn Gly Asn Gly 145 150 160

Asn Met Thr Arg Thr Leu Leu Thr Ser Gly Leu Ser Asn Asp Gly Val

Ser Thr Thr Gly Phe Arg Ser Ala Glu Ala Leu Phe Glu Lys Ala Val

Thr Pro Ser Asp Val Gly Lys Leu Asn Arg Leu Val Ile Pro Lys His 195 200 205

His Ala Glu Lys His Phe Pro Leu Pro Ser Ser Asn Val Ser Val Lys 210 215 220

Gly Val Leu Leu Asn Phe Glu Asp Val Asn Gly Lys Val Trp Arg Phe 225 230 235 240

Arg Tyr Ser Tyr Trp Asn Ser Ser Gln Ser Tyr Val Leu Thr Lys Gly 245 250 255

Trp Ser Arg Phe Val Lys Glu Lys Asn Leu Arg Ala Gly Asp Val Val 260 265 270

Ser Phe Ser Arg Ser Asn Gly Gln Asp Gln Gln Leu Tyr Ile Gly Trp 275 280 285

Lys Ser Arg Ser Gly Ser Asp Leu Asp Ala Gly Arg Val Leu Arg Leu 290 295 300

Phe Gly Val Asn Ile Ser Pro Glu Ser Ser Arg Asn Asp Val Val Gly 305 310 315 320

### mbi19 Sequence Listing.ST25

Asn Lys Arg Val Asn Asp Thr Glu Met Leu Ser Leu Val Cys Ser Lys 325 330 335

Lys Gln Arg Ile Phe His Ala Ser

<211> 724 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (20)..(694) <223> G912 <400> 23 catcttatcc aaagaaaaa atg aat cca ttt tac tct aca ttc cca gac tcg
Met Asn Pro Phe Tyr Ser Thr Phe Pro Asp Ser

1 5 10 52 ttt ctc tca atc tcc gat cat aga tct ccg gtt tca gac agt agt gag Phe Leu Ser Ile Ser Asp His Arg Ser Pro Val Ser Asp Ser Ser Glu 15 20 25100 148 tgt tca cca aag tta gct tca agt tgt cca aag aaa cga gct ggg agg Cys Ser Pro Lys Leu Ala Ser Ser Cys Pro Lys Lys Arg Ala Gly Arg aag aag ttt cgt gag aca cgt cat ccg att tac aga gga gtt cgt cag Lys Lys Phe Arg Glu Thr Arg His Pro Ile Tyr Arg Gly Val Arg Gln 45 50 55 196 agg aat tot ggt aaa tgg gtt tgt gaa gtt aga gag cot aat aag aaa Arg Asn Ser Gly Lys Trp Val Cys Glu Val Arg Glu Pro Asn Lys Lys 244 tct agg att tgg tta ggt act ttt ccg acg gtt gaa atg gct gct cgt Ser Arg Ile Trp Leu Gly Thr Phe Pro Thr Val Glu Met Ala Ala Arg 292 gct cat gat gtt gct gct tta gct ctt cgt ggt cgc tct gct tgt ctc Ala His Asp Val Ala Ala Leu Ala Leu Arg Gly Arg Ser Ala Cys Leu 340 aat ttc gct gat tct gct tgg cgg ctt cgt att cct gag act act tgt Asn Phe Ala Asp Ser Ala Trp Arg Leu Arg Ile Pro Glu Thr Thr Cys 110 115 120 388 cct aag gag att cag aaa gct gcg tct gaa gct gca atg gcg ttt cag Pro Lys Glu Ile Gln Lys Ala Ala Ser Glu Ala Ala Met Ala Phe Gln 436 aat gag act acg acg gag gga tct aaa act gcg gcg gag gca gag gag Asn Glu Thr Thr Glu Gly Ser Lys Thr Ala Ala Glu Ala Glu Glu 484 532 aat ggt ggt gtg ttt tat atg gat gat gag gcg ctt ttg ggg atg ccc 580 Asn Gly Gly Val Phe Tyr Met Asp Asp Glu Ala Leu Leu Gly Met Pro aac ttt ttt gag aat atg gcg gag ggg atg ctt ttg ccg ccg ccg gaa Asn Phe Phe Glu Asn Met Ala Glu Gly Met Leu Leu Pro Pro Pro Glu 628 gtt ggc tgg aat cat aac gac ttt gac gga gtg ggt gac gtg tca ctc 676

mbi19 Sequence Listing.ST25 Val Gly Trp Asn His Asn Asp Phe Asp Gly Val Gly Asp Val Ser Leu 205 210 215

tgg agt ttt gac gag taa ttttttggct ctttttctgg ataataagtt Trp Ser Phe Asp Glu

724

<210> 24 <211> 224 <212> PRT

<213> Arabidopsis thaliana

<400> 24

Met Asn Pro Phe Tyr Ser Thr Phe Pro Asp Ser Phe Leu Ser Ile Ser 1 10 15

Asp His Arg Ser Pro Val Ser Asp Ser Ser Glu Cys Ser Pro Lys Leu 20 30

Ala Ser Ser Cys Pro Lys Lys Arg Ala Gly Arg Lys Lys Phe Arg Glu 35 40 45

Thr Arg His Pro Ile Tyr Arg Gly Val Arg Gln Arg Asn Ser Gly Lys 50 55 60

Trp Val Cys Glu Val Arg Glu Pro Asn Lys Lys Ser Arg Ile Trp Leu

Gly Thr Phe Pro Thr Val Glu Met Ala Ala Arg Ala His Asp Val Ala 85 90 95

Ala Leu Ala Leu Arg Gly Arg Ser Ala Cys Leu Asn Phe Ala Asp Ser 100 105 110

Ala Trp Arg Leu Arg Ile Pro Glu Thr Thr Cys Pro Lys Glu Ile Gln
115 120 125

Lys Ala Ala Ser Glu Ala Ala Met Ala Phe Gln Asn Glu Thr Thr Thr

Glu Gly Ser Lys Thr Ala Ala Glu Ala Glu Glu Ala Ala Gly Glu Gly

Val Arg Glu Gly Glu Arg Arg Ala Glu Glu Gln Asn Gly Gly Val Phe

Tyr Met Asp Asp Glu Ala Leu Leu Gly Met Pro Asn Phe Phe Glu Asn

Met Ala Glu Gly Met Leu Leu Pro Pro Pro Glu Val Gly Trp Asn His

Asn Asp Phe Asp Gly Val Gly Asp Val Ser Leu Trp Ser Phe Asp Glu 210 215 220

<210> 25 <211> 1082 <212> DNA

<213> Arabidopsis thaliana

### mbil9 Sequence Listing.ST25

<220>

CDS <221> <222> (53)..(1063) G996 <400> 25 cgatcgatct tgaattgatt ctttgtagta ttttatttac atatatatat ag atg gga 58 aga cat tca tgt tgt tac aaa cag aaa ctg agg aaa gga ctt tgg tct Arg His Ser Cys Cys Tyr Lys Gln Lys Leu Arg Lys Gly Leu Trp Ser 106 cct gaa gaa gat gag aag ctt ctt cgt tac atc act aag tat ggt cat Pro Glu Glu Asp Glu Lys Leu Leu Arg Tyr Ile Thr Lys Tyr Gly His 20 25 30154 ggt tgc tgg agc tct gtc cct aaa caa gct ggt tta cag aga tgt gga Gly Cys Trp Ser Ser Val Pro Lys Gln Ala Gly Leu Gln Arg Cys Gly 202 250 aga gga gca ttt tct caa gat gaa gaa aat ctc att att gaa ctt cat Arg Gly Ala Phe Ser Gln Asp Glu Glu Asn Leu Ile Ile Glu Leu His 298 gcc gtt ctt ggc aat aga tgg tct cag ata gct gca cag ctt cct gga Ala Val Leu Gly Asn Arg Trp Ser Gln Ile Ala Ala Gln Leu Pro Gly 346 aga acc gac aat gaa atc aag aat ctt tgg aat tct tgt ttg aag aag Arg Thr Asp Asn Glu Ile Lys Asn Leu Trp Asn Ser Cys Leu Lys Lys aaa ttg agg ctg aga gga att gac ccg gtt aca cac aag ctc tta acc Lys Leu Arg Leu Arg Gly Ile Asp Pro Val Thr His Lys Leu Leu Thr 442 gaa atc gaa acc ggt aca gat gac aaa aca aaa ccg gtt gag aag agt Glu Ile Glu Thr Gly Thr Asp Asp Lys Thr Lys Pro Val Glu Lys Ser 135 140 145490 caa cag acc tac ctc gtt gag act gat ggc tcc tct agt acc act act Gln Gln Thr Tyr Leu Val Glu Thr Asp Gly Ser Ser Ser Thr Thr Thr 538 tgt agt act aac caa aac aac act gat cat ctt tat acc gga aat Cys Ser Thr Asn Gln Asn Asn Asn Thr Asp His Leu Tyr Thr Gly Asn 586 ttc ggt ttt caa cgg tta agt cta gaa aac ggt tca aga atc gca gcc Phe Gly Phe Gln Arg Leu Ser Leu Glu Asn Gly Ser Arg Ile Ala Ala 634 ggt tct gac ctc ggt atc tgg att ccc caa acc gga aga aac cat cat Gly Ser Asp Leu Gly Ile Trp Ile Pro Gln Thr Gly Arg Asn His His 682 cat cat gtc gat gaa acc atc cct agt gca gtg gta cta ccc ggt tca His His Val Asp Glu Thr Ile Pro Ser Ala Val Val Leu Pro Gly Ser 730 atg ttc tca tcc ggt tta acc ggt tat aga tcc tcc aat ctc ggt tta Met Phe Ser Ser Gly Leu Thr Gly Tyr Arg Ser Ser Asn Leu Gly Leu 778 att gaa ttg gaa aac tca ttc tca acc ggg cca atg atg aca gag cat Ile Glu Leu Glu Asn Ser Phe Ser Thr Gly Pro Met Met Thr Glu His 826

mbil9	Sequence	Ticking	CTOE
IIID T T 3	sequence	Listing	5123

		mbil9 Seque	nce Listing.ST25	
		n Tyr Asn Asn	tca aca ttc ttt Ser Thr Phe Phe 270	
			gaa aat caa aat Glu Asn Gln Asn 285	
			tac agt gat ata Tyr Ser Asp Ile	
			aat gtt ggt atg Asn Val Gly Met 320	
	Leu Gln Pro Gl		tat ggc cat ata Tyr Gly His Ile 335	taa 1063
atcttcttgt	atattataa			1082
<210> 26 <211> 336 <212> PRT <213> Arab	oidopsis thalian	a		
<400> 26				
Met Gly Arg	His Ser Cys Cy 5	s Tyr Lys Gln 10	Lys Leu Arg Lys	Gly Leu 15
Trp Ser Pro	Glu Glu Asp Gl 20	u Lys Leu Leu 25	Arg Tyr Ile Thr 30	Lys Tyr
GÍy His Gly 35	Cys Trp Ser Se	r Val Pro Lys 40	Gln Ala Gly Leu 45	Gln Arg
Cys Gly Lys	Ser Cys Arg Le 55	u Arg Trp Ile	Asn Tyr Leu Arg 60	Pro Asp

Leu Lys Arg Gly Ala Phe Ser Gln Asp Glu Glu Asn Leu Ile Ile Glu 65 70 75 80

Leu His Ala Val Leu Gly Asn Arg Trp Ser Gln Ile Ala Ala Gln Leu 85 90 95

Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn Leu Trp Asn Ser Cys Leu 100 105 110

Lys Lys Leu Arg Leu Arg Gly Ile Asp Pro Val Thr His Lys Leu  $\cdot\,$  115  $\,$  120  $\,$  125

Leu Thr Glu Ile Glu Thr Gly Thr Asp Asp Lys Thr Lys Pro Val Glu 130 135

Lys Ser Gln Gln Thr Tyr Leu Val Glu Thr Asp Gly Ser Ser Ser Thr 145 150 160

Thr Thr Cys Ser Thr Asn Gln Asn Asn Asn Thr Asp His Leu Tyr Thr 165 170 175

mbi19 Sequence Listing.ST25 Gly Asn Phe Gly Phe Gln Arg Leu Ser Leu Glu Asn Gly Ser Arg Ile Ala Ala Gly Ser Asp Leu Gly Ile Trp Ile Pro Gln Thr Gly Arg Asn His His His His Val Asp Glu Thr Ile Pro Ser Ala Val Val Leu Pro Gly Ser Met Phe Ser Ser Gly Leu Thr Gly Tyr Arg Ser Ser Asn Leu 225 230 240 Gly Leu Ile Glu Leu Glu Asn Ser Phe Ser Thr Gly Pro Met Met Thr Glu His Gln Gln Ile Gln Glu Ser Asn Tyr Asn Asn Ser Thr Phe Phe Gly Asn Gly Asn Leu Asn Trp Gly Leu Thr Met Glu Glu Asn Gln Asn 275 280 285Pro Phe Thr Ile Ser Asn His Ser Asn Ser Ser Leu Tyr Ser Asp Ile 295 Lys Ser Glu Thr Asn Phe Phe Gly Thr Glu Ala Thr Asn Val Gly Met Trp Pro Cys Asn Gln Leu Gln Pro Gln Gln His Ala Tyr Gly His Ile 330 <210> 27 <211> 1606 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (150)..(1310) <223> G1068 <400> 27 gagagttgtt agctagctca cacgettteg ettaaaacte aaaaacetge actttetegt 60 120 ctattttctc ggcattcgta aaacagaaaa gtgggtctcc aagaaaatta ccctaaattc acaaagattc atacttttct ccacctcca atg gat tcc aga gag atc cac cac Met Asp Ser Arg Glu Ile His His 173 221 caa cat cta caa caa cag caa caa cca ccg cca ggg atg tta atg agt Gln His Leu Gln Gln Gln Gln Pro Pro Pro Gly Met Leu Met Ser 269 cac cac aat tcc tac aat cga aac cct aac gcc gcc gcc gct gtt tta His His Asn Ser Tyr Asn Arg Asn Pro Asn Ala Ala Ala Ala Leu 317 atg ggt cac aac acc tcc aca tct caa gct atg cat caa aga tta cct Met Gly His Asn Thr Ser Thr Ser Gln Ala Met His Gln Arg Leu Pro 365 65

# mbi19 Sequence Listing.ST25

ttt Phe	ggt Gly	ggt Gly 75	tct Ser	atg Met	tca Ser	ccg Pro	cat His 80	cag Gln	cct Pro	caa Gln	caa Gln	cat His 85	cag Gln	tat Tyr	cat His	413
cat Kis	cct Pro 90	cag Gln	cct Pro	cag Gln	caa Gln	cag Gln 95	ata Ile	gat Asp	cag Gln	aag Lys	act Thr 100	ctt Leu	gaa Glu	tct Ser	ctt Leu	461
gga Gly 105	ttt Phe	cct Pro	act Thr	tcg Ser	cct Pro 110	ctt Leu	cct Pro	tct Ser	gct Ala	tct Ser 115	aat Asn	tct Ser	tac Tyr	ggt Gly	ggt Gly 120	509
gga Gly	aat Asn	gaa Glu	gga Gly	ggt Gly 125	ggt Gly	ggt Gly	ggt Gly	gat Asp	agc Ser 130	gcc Ala	gga Gly	gct Ala	aat Asn	gct Ala 135	aac Asn	557
tct Ser	tcc Ser	gat Asp	cca Pro 140	cct Pro	gct Ala	aaa Lys	cgg Arg	aac Asn 145	aga Arg	gga Gly	cgt Arg	cct Pro	cct Pro 150	ggc Gly	tcc Ser	605
ggt Gly	aag Lys	aag Lys 155	cag Gln	ctc Leu	gat Asp	gct Ala	tta Leu 160	gga Gly	gga Gly	aca Thr	gga Gly	gga Gly 165	gtt Val	ggg Gly	ttc Phe	653
acg Thr	cct Pro 170	cat His	gtc Val	att Ile	gag Glu	gtt Val 175	aaa Lys	aca Thr	gga Gly	gag Glu	gac Asp 180	ata Ile	gct Ala	acg Thr	aag Lys	701
ata Ile 185	ttg Leu	gcg Ala	ttt Phe	acg Thr	aac Asn 190	caa Gln	999 Gly	cca Pro	cgc Arg	gca Ala 195	atc Ile	tgt Cys	att Ile	ctc Leu	tca Ser 200	749
gct Ala	aca Thr	gga Gly	gct Ala	gta Val 205	act Thr	aat Asn	gtg Val	atg Met	ctt Leu 210	cgt Arg	caa Gln	gct Ala	aac Asn	aat Asn 215	agc Ser	797
aat Asn	cct Pro	act Thr	gga Gly 220	act Thr	gtt Val	aag Lys	tat Tyr	gag Glu 225	ggc Gly	cga Arg	ttt Phe	gaa Glu	atc Ile 230	att Ile	tct Ser	845
ctg Leu	tca Ser	ggt Gly 235	tct Ser	ttc Phe	ttg Leu	aat Asn	tct Ser 240	gag Glu	agt Ser	aat Asn	ggt Gly	act Thr 245	gtg Val	acc Thr	aaa Lys	893
act Thr	ggt Gly 250	aac Asn	ttg Leu	agt Ser	gtg Val	tcg Ser 255	ctg Leu	gct Ala	gga Gly	cac His	gaa Glu 260	ggc Gly	cgg Arg	att Ile	gtg Val	941
ggt Gly 265	gga Gly	tgt Cys	gtt Val	gat Asp	gga Gly 270	atg Met	cta Leu	gta Val	gct Ala	gga Gly 275	tca Ser	caa Gln	gtc Val	cag Gln	gtc Val 280	989
att Ile	gtg Val	gga Gly	agc Ser	ttt Phe 285	gta Val	cca Pro	gat Asp	gga Gly	agg Arg 290	aag Lys	cag Gln	aaa Lys	caa Gln	agt Ser 295	gcg Ala	1037
999 Gly	cgt Arg	gct Ala	cag Gln 300	aat Asn	act Thr	ccg Pro	gag Glu	cca Pro 305	gct Ala	tca Ser	gca Ala	cca Pro	gcc Ala 310	aat Asn	atg Met	1085
						ggt Gly										1133
						tca Ser 335										1181
						aac Asn										1229
						ccg Pro			Gln		Pro					1277

	mbil9	Sequence	Listing.ST25	
365		370		375

cag aat ctc tgg cct ggc aac agt cct caa taa acagatggtt catgggtcaa Gln Asn Leu Trp Pro Gly Asn Ser Pro Gln 1330 1390 gatttgaceg ggtttgcttc tctgttcctt ttgacacatc tctccatcag atttatctct 1450 ataaagtaga ttgagctctc ttactctctc atcttcttct cctttactat ttctcttaaa tttagctttg gttttagata aatagagaga gagagacatg ttaagtaggt ttcaaattca 1510 atcttgttta gtttgtttct tagtagtttc ttttgattgt gatgatcata aagacttgtt 1570 1606 ctttttctcc tatattcaac gaattatcca ctttaa

<210> 28 <211> 386 <212> PRT

<213> Arabidopsis thaliana

Met Asp Ser Arg Glu Ile His His Gln Gln Gln Gln Gln Gln Gln I 10 15

Gln Gln Gln Gln Gln Gln Gln His Leu Gln Gln Gln Gln Gln

Pro Pro Gly Met Leu Met Ser His His Asn Ser Tyr Asn Arg Asn 35 40 45

Pro Asn Ala Ala Ala Ala Val Leu Met Gly His Asn Thr Ser Thr Ser 50 60

Gln Ala Met His Gln Arg Leu Pro Phe Gly Gly Ser Met Ser Pro His 65 70 75 80

Gln Pro Gln Gln His Gln Tyr His His Pro Gln Pro Gln Gln Gln Ile 85 90 95

Asp Gln Lys Thr Leu Glu Ser Leu Gly Phe Pro Thr Ser Pro Leu Pro

Ser Ala Ser Asn Ser Tyr Gly Gly Gly Asn Glu Gly Gly Gly Gly 115 120 125

Asp Ser Ala Gly Ala Asn Ala Asn Ser Ser Asp Pro Pro Ala Lys Arg

Asn Arg Gly Arg Pro Pro Gly Ser Gly Lys Lys Gln Leu Asp Ala Leu 145 150 155 160

Gly Gly Thr Gly Gly Val Gly Phe Thr Pro His Val Ile Glu Val Lys

Thr Gly Glu Asp Ile Ala Thr Lys Ile Leu Ala Phe Thr Asn Gln Gly 180 180 190

Pro Arg Ala Ile Cys Ile Leu Ser Ala Thr Gly Ala Val Thr Asn Val

mbil9 Sequence Listing.ST25

	mbily Sequence	Listing. S125	
Met Leu Arg Gln Ala Asn As 210 21		Gly Thr Val Lys 220	: Tyr
Glu Gly Arg Phe Glu Ile Il 225 230	e Ser Leu Ser Gly 235		1 Ser 240
Glu Ser Asn Gly Thr Val Th	r Lys Thr Gly Asn 250	Leu Ser Val Ser 255	Leu
Ala Gly His Glu Gly Arg Il	e Val Gly Gly Cys 265	Val Asp Gly Met 270	Leu
Val Ala Gly Ser Gln Val Gl 275	n Val Ile Val Gly 280	Ser Phe Val Pro 285	Asp
Gly Arg Lys Gln Lys Gln Se 290 29		Gln Asn Thr Pro	Glu
Pro Ala Ser Ala Pro Ala As 305 310	n Met Leu Ser Phe 315		Gly 320
Pro Gly Ser Pro Arg Ser Gl 325	n Gly Gln Gln His 330	Ser Ser Glu Ser 335	Ser
Glu Glu Asn Glu Ser Asn Se 340	r Pro Leu His Arg 345	Arg Ser Asn Asr 350	Asn
Asn Ser Asn Asn His Gly II 355	e Phe Gly Asn Ser 360	Thr Pro Gln Pro 365	Leu
His Gln Ile Pro Met Gln Me 370 37		Trp Pro Gly Asr	Ser
Pro Gln 385			
<210> 29 <211> 1630 <212> DNA <213> Arabidopsis thalian	a		
<220> <221> CDS <222> (97)(1398) <223> G1337			
<400> 29 aatggatttg tcatcattct tctc	accgtc cttagtctct	gaaaataaat toto	attttg 60
atttegaatt ttagggattt tgag	agagag tcagtt atg Met 1	agt agt tcg gag Ser Ser Ser Gli 5	aga 114 Arg
gta ccg tgc gat ttc tgc gg Val Pro Cys Asp Phe Cys Gl 10	c gag cgt acg gcg y Glu Arg Thr Ala 15	gtt ttg ttt tgt Val Leu Phe Cys 20	aga 162 Arg
gcc gat acg gcg aag ctg tg Ala Asp Thr Ala Lys Leu Cy 25	t ttg cct tgt gat s Leu Pro Cys Asp 30	cag caa gtt cac Gln Gln Val His 35	acg 210 Thr

gcg aat ctg ttg dag agg aag cac gtg cga tct cgg atc tgc gat aat la Aan Leu Leu Ser Arg Lya His Val Arg Ser Gin Ite Cys Asp Aan         258           tgc ggt aac gag cca gtc gat ctg gt ttg gt gt ttg gg tt ttg gag aat ctg gag gag gag gag gag gag gag gag gag g												List					
Cys         Gly         Asn         Clu         Pro         Val         Ser         Val         Arg         Cys         Fib         Thr         Asp         Ash         Leu         To           teg         ag         teg         teg         teg         teg         teg         teg         ag         teg         teg         ag         teg	Ala	a Asn	ctg Leu	ttg Leu	tcg Ser	agg Arg	Lys	cac His	gtg Val	cga Arg	tct Ser	Gln	atc Ile	tgc Cys	gat Asp	aat Asn	258
Leu Cys Gin Giu Cys Asp Trp Asp Val Bis Gly Ser Cys Ser Val Ser  gat gct cat gtt cga tcc gcc gtg gaa ggt ttt tcc ggt tgt cca tcg Asp Ala His Val Arg Ser Ala Val Glu Gly Phe Ser Gly Cys Pro Ser  gcg ttg gag ctt gct gct gct gt gag gag ggt ttt tcc ggt tgt cca tcg Ala Leu Glu Leu Ala Ala Leu Trp Gly Leu Asp Leu Glu Gln Gly Arg  gcg ttg gag aga aat caa gtt ccg atg atg gcg atg atg atg gga aga gat Lys Asp Glu Glu Asn Gln Val Pro Met Met Ala Met Met Asp Asn 120  ttc ggg atg cag ttg gat tct tgg gtt ttg ggg atct aat gaa ttg atg He Gly Met Gln Leu Asp Ser Trp Val Leu Gly Ser Asn Glu Leu Ile 135  ttc ggg atg cag ttg gat tct ttg gg atct ttg gga tct aat gaa ttg Asp Shr Trp Trh Trh Phe Lys Lys Arg Gly Ser Cys Gly Ser Ser 155  tgt ggg agg tat aag cag gta tg	Су	ggt Gly	aac Asn	gag Glu	cca Pro	Val	tct Ser	gtt Val	cgg Arg	tgt Cys	Phe	acc Thr	gat Asp	aat Asn	ctg Leu	He	306
Asp Ala His Val Arg Ser Ala Val Glu Gly Phe Ser Gly Cys Pro Ser Glu Glu Gly Phe Ser Gly Cys Pro Ser Glu	Lei	g tgt i Cys	cag Gln	gag Glu	Сув	gat Asp	tgg Trp	gat Asp	gtt Val	His	gga Gly	agt Ser	tgt Cys	tca Ser	Val	tcc Ser	354
Ala Leu Glu Leu Ala Ala Leu Trp Gly Leu Asp Leu Glu Glu Gly Arg  aaa gat gaa gag aat caa gtt ccg atg atg acg atg atg atg atg atg atg atg atg atg at	ga As	gct Ala	cat His	Val	cga Arg	tcc Ser	gcc Ala	gtg Val	Glu	ggt Gly	ttt Phe	tcc Ser	ggt Gly	Cys	cca Pro	tcg Ser	402
Lys   Asp   Glu   Glu   Asn   Glu   Yal   Pro   Met   Met   Ala   Met   Met   Asp   Asn   Asn   125	gc:	g ttg a Leu	Glu	ctt Leu	gct Ala	gct Ala	tta Leu	Trp	gga Gly	ctt Leu	gat Asp	ttg Leu	Glu	caa Gln	999 Gly	agg Arg	450
## Phe Gly Met Gln Leu Asp Ser Trp Val Leu Gly Ser Asn Glu Leu 116  gtt ccc agc gat acg acg ttt aag aag cgt gga tct tgt gga tct agt agg ggg agg agg act as ag cag ttg tgt aag gag ggt gtt Leu Cys Lys Asp Gly Asp Arg Tyr Lys Gln Val Leu Cys Lys Gn Leu Glu Glu Leu Leu Leu 175  gat ggg agg tgt gtt gtc ggt ggt gat ggc gat ggt gat ggt gat ggt gat ggt gat cgt gac cgt Lys Ser Gly Val Val Gly Gly Asp Gly Asp Asp Gly Asp Arg Asp Arg Asp Arg 190  gat tgt gac cgt gag ggt gct tgt gat ggt gat ga	aaa Lys	a Asp	gaa Glu	gag Glu	aat Asn	caa Gln	Val	ccg Pro	atg Met	atg Met	gcg Ala	Met	atg Met	atg Met	gat Asp	aat Asn	498
Val         Pro         Ser         Asp         Thr         Thr         Phe         Lys         Lys         Arg         Gly         Ser         Cys         Gly         Ser         Gly         Lys         Gln         Val         Lys         Lys         Gln         Leu         Cys         Lys         Gln         Leu         Cys         Lys         Gln         Leu         Gly         Leu         Leu <td>Phe</td> <td>e Gly</td> <td>atg Met</td> <td>cag Gln</td> <td>t tg Leu</td> <td>Asp</td> <td>tct Ser</td> <td>tgg Trp</td> <td>gtt Val</td> <td>ttg Leu</td> <td>Gly</td> <td>tct Ser</td> <td>aat Asn</td> <td>gaa Glu</td> <td>ttg Leu</td> <td>Ile</td> <td>546</td>	Phe	e Gly	atg Met	cag Gln	t tg Leu	Asp	tct Ser	tgg Trp	gtt Val	ttg Leu	Gly	tct Ser	aat Asn	gaa Glu	ttg Leu	Ile	546
Cys         Gly         Arg         Tyr         Lys         Gln         Val         Leu         Cys         Lys         Gln         Leu         Leu         Leu         Gly         Lys         Glu         Glu         Leu         Gly         Gly         Gly         Asp         Gly         Asp         Asp         Gly         Asp         Arg         Gly         Asp         Asp         Gly         Asp         Arg         Gly         Ala         Cys         Asp         Gly         Asp         Arg         Gly         Ala         Cys         Asp         Gly         Asp         Arg         Gly         Ala         Cys         Asp         Gly         Asp         Arg         Arg <td>gt: Va:</td> <td>ccc Pro</td> <td>agc Ser</td> <td>gat Asp</td> <td>Thr</td> <td>acg Thr</td> <td>ttt Phe</td> <td>aag Lys</td> <td>aag Lys</td> <td>Arg</td> <td>gga Gly</td> <td>tct Ser</td> <td>tgt Cys</td> <td>gga Gly</td> <td>Ser</td> <td>agt Ser</td> <td>594</td>	gt: Va:	ccc Pro	agc Ser	gat Asp	Thr	acg Thr	ttt Phe	aag Lys	aag Lys	Arg	gga Gly	tct Ser	tgt Cys	gga Gly	Ser	agt Ser	594
Lys         Ser         Gly         Val         Val         Gly         Asp         Gly         Asp         Gly         Asp         Asp <td>tg: Cy:</td> <td>ggg Gly</td> <td>agg Arg</td> <td>Tyr</td> <td>aag Lys</td> <td>cag Gln</td> <td>gta Val</td> <td>ttg Leu</td> <td>Cys</td> <td>aag Lys</td> <td>cag Gln</td> <td>ctt Leu</td> <td>gag Glu</td> <td>Glu</td> <td>ttg Leu</td> <td>ctt Leu</td> <td>642</td>	tg: Cy:	ggg Gly	agg Arg	Tyr	aag Lys	cag Gln	gta Val	ttg Leu	Cys	aag Lys	cag Gln	ctt Leu	gag Glu	Glu	ttg Leu	ctt Leu	642
Asp Cys Asp Arg Glu Gly Ala Cys Asp Gly Asp Gly Asp Gly Asp Gly Glu Ala  gga gag gag ggg ctt at ggt ccg gag atg tca gag aga atg tca gag aga ttg aaa tgg tca  reflection of the company of	aaq Ly:	g agt s Ser	Gly	gtt Val	gtc Val	ggt Gly	ggt Gly	Asp	ggc Gly	gat Asp	gat Asp	ggt Gly	Asp	cgt Arg	gac Asp	cgt Arg	690
aga gat gtt gag gag att act act act act act act gag gag gga ggt ggy ggy ggy ggy ggy ggy ggy ggy ggy	ga! Asj	Cys	Asp Asp	cgt Arg	gag Glu	ggt Gly	Ala	tgt Cys	gat Asp	gga Gly	gat Asp	Gly	gat Asp	gga Gly	gaa Glu	gca Ala	738
Arg Asp Val Glu Glu Ile Asn Gly Gly Gly Gly Gly Gly Val Asn Gln 235  Cag tgg aat ggt act act act act act act cat agt ggt ggc cag agt tct cag Gln Trp Asn Ala Thr Thr Thr Asn Pro Ser Gly Gly Gln Ser Ser Gln 255  ata tgg gat ttt aac ttg gga cag tca cgg gga cct gag gat acg agt Thr Ser 275  Cag gtg gaa gct gca tat gta ggg aaa ggt ggc Gly Asp Thr Ser 275  Cag gtg gaa gct gca tat gta ggg aaa ggt Lys Gly Lys Gly Ala Ala Ser Ser Phe Thr 280  atc aac aat ttt gtt gac cat atg aat gaa act tgt tcc act act act act act asg aat ggt Thr Asn Asn Asn Phe Val Asp His Met Asn Glu Thr Cys Ser Thr Asn Val 295  aaa ggt gtc aaa gag att aaa aag gag att aaa aag gat gac tac aag cga tca acg act tca tca act tca tca tca act tca Lys Gly Val Lys Gly Val Lys Gly Ile Lys Lys Asp Asp Tyr Lys Arg Ser Thr Ser 325  ggc cag gta caa cca aca aaa tct gag agc aac aaa tct gag aac act act act act act acc acc acc acc	Gl	/ Glu	ggg Gly	ctt Leu	atg Met	Val	ccg Pro	gag Glu	atg Met	tca Ser	Glu	aga Arg	ttg Leu	aaa Lys	tgg Trp	Ser	786
At a tgg gat ttt aac ttg gga cag tca cgg gga cct gag gat acg agt acg agt Trp Asn Phe Asn Leu Gly Gln Ser Arg Gly Pro Glu Asp Thr Ser 275  Cga gtg gaa gct gca tat gta ggg aaa ggt Lys Gly Lys Gly Ala Ala Ser Ser Phe Thr 280  atc aac aat ttt gtt gac cat atg aat gaa gat 280 Ser Asn Glu Thr Cys Ser Thr Asn Val 295  aaa ggt gtc aaa gag att aaa aag gat gac tac aag ggt gtc tca act atg gtg 1026  Lys Gly Val Lys Gly Val Lys Gly Lys Asp Asp Tyr Lys Asp Ser Thr Ser 325  ggc cag gta caa cca aca act aaa tct gag agc acc as aag agc acc acc acc acc acc acc acc acc	aga Arg	a gat g Asp	gtt Val	gag Glu •	Glu	atc Ile	aat Asn	ggt Gly	ggc Gly	Gly	gga Gly	gga Gly	gga Gly	gtt Val	Asn	cag Gln	834
The Trp Asp Phe Asn Leu Gly Gln Ser Arg Gly Pro Glu Asp Thr Ser 275  Cga gtg gaa gct gca tat gta ggg aaa ggt gct gct tct tca ttc aca 978  Arg Val Glu Ala Ala Tyr Val Gly Lys Gly Ala Ala Ser Ser Phe Thr 280  atc aac aat ttt gtt gac cat atg aat gaa act tgt tcc act aat gtg 1026  Ile Asn Asn Phe Val Asp His Met Asn Glu Thr Cys Ser Thr Asn Val 310  aaa ggt gtc aaa gag att aaa aag gat gac tac aag cga tca act tca Lys Gly Val Lys Gly Val Lys Asp Asp Asp Tyr Lys Arg Ser Thr Ser 325  ggc cag gta caa cca aca aaa tct gag agc aac aat cgt cca att acc Gly Gln Val Gln Pro Thr Lys Ser Glu Ser Asn Asn Arg Pro Ile Thr 330  1122	Gli	g tgg n Trp	aat Asn	Ala	act Thr	act Thr	act Thr	aat Asn	Pro	agt Ser	ggt Gly	ggc Gly	cag Gln	Ser	tct Ser	cag Gln	882
Arg Val Glu Ala Ala Tyr Val Gly Lys Gly Ala Ala Ala Ser Ser Phe Thr  atc aac aat ttt gtt gac cat atg aat gaa act tgt tcc act aat gtg Ile Asn Asn Phe Val Asp His Met Asn Glu Thr Cys Ser Thr Asn Val 300  aaa ggt gtc aaa gag att aaa aag gat gac tac aag cga tca act tca Lys Gly Val Lys Glu Ile Lys Lys Asp Asp Tyr Lys Arg Ser Thr Ser 315  ggc cag gta caa cca aca aaa tct gag agc aac aat cgt cca att acc Gly Gln Val Gln Pro Thr Lys Ser Glu Ser Asn Asn Arg Pro Ile Thr 330  1026  1026  1026  1026  1027  1028  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029  1029	ata Ile	tgg Trp	Asp	ttt Phe	aac Asn	ttg Leu	gga Gly	Gln	tca Ser	cgg Arg	gga Gly	cct Pro	Glu	gat Asp	acg Thr	agt Ser	930
Ile Asn Asn Phe Val Asp His Met Asn Glu Thr Cys Ser Thr Asn Val 300 310  aaa ggt gtc aaa gag att aaa aag gat gac tac aag cga tca act tca Lys Gly Val Lys Glu Ile Lys Lys Asp Asp Tyr Lys Arg Ser Thr Ser 315 310  ggc cag gta caa cca aca aaa tct gag agc aac aat cgt cca att acc Gly Gln Val Gln Pro Thr Lys Ser Glu Ser Asn Asn Arg Pro Ile Thr 330 1122	cga Arg	y Val	gaa Glu	gct Ala	gca Ala	tat Tyr	Val	999 Gly	aaa Lys	ggt Gly	gct Ala	Ala	tct Ser	tca Ser	ttc Phe	aca Thr	978
Lys Gly Val Lys Glu Ile Lys Lys Asp Asp Tyr Lys Arg Ser Thr Ser 315 320 325  ggc cag gta caa cca aca aaa tct gag agc aac aat cgt cca att acc Gly Gln Val Gln Pro Thr Lys Ser Glu Ser Asn Asn Arg Pro Ile Thr 330 335 340	110	e Asn				Asp					Thr					Val	1026
Gly Gln Val Gln Pro Thr Lys Ser Glu Ser Asn Asn Arg Pro Ile Thr 330 335 340	aaa Lys	ggt Gly	gtc Val	aaa Lys	Glu	att Ile	aaa Lys	aag Lys	gat Asp	Asp	tac Tyr	aag Lys	cga Arg	tca Ser	Thr	tca Ser	1074
Dama 3.0	gg( Gl	cag Gln	gta Val	Gln	cca Pro	aca Thr	aaa Lys	tct Ser	Glu	Ser	Asn	Asn	cgt Arg	Pro	att Ile	acc Thr	1122

mbilo	Sequence	* : :	CT2 F
mbile	Sequence	Listing	. 5125

ttt ggc tct gag aaa ggt tcg Phe Gly Ser Glu Lys Gly Ser 345			1170
gag cat att gct gga act agt Glu His Ile Ala Gly Thr Ser 360 365			1218
aag gct gat ctg gag cgg ctg Lys Ala Asp Leu Glu Arg Leu 375 380			1266
cgt tac aag gaa aag agg aag Arg Tyr Lys Glu Lys Arg Lys 395			1314
tat gaa tcg agg aag gca aga Tyr Glu Ser Arg Lys Ala Arg 410	gct gac act agg Ala Asp Thr Arg 415	ttg cgt gtc aga ggc Leu Arg Val Arg Gly 420	1362
aga ttt gtg aaa gct agt gaa Arg Phe Val Lys Ala Ser Glu 425		taa ccttaagttt	1408
tttcacatag gcttcctttt agctac	aaac ttagttactt	tttttactcc actgcctcat	1468
aaatgtacag accggtctcg tttcat	ctgg ccgcccttct	tgttttattg ccttatctgg	1528
cccttttatg taccttggaa tcttat	ctag tttaaaaaag	attgtaacct tctagaaaac	1588
catattetgt tgacagtata tacatg	tcta tccaagcaaa	aa	1630
-210> 30			

<400> 30

Met Ser Ser Ser Glu Arg Val Pro Cys Asp Phe Cys Gly Glu Arg Thr 1  $\phantom{\bigg|}10\phantom{\bigg|}$  15

Ala Val Leu Phe Cys Arg Ala Asp Thr Ala Lys Leu Cys Leu Pro Cys 20 30

Asp Gln Gln Val His Thr Ala Asn Leu Leu Ser Arg Lys His Val Arg 35 40 45

Ser Gln Ile Cys Asp Asn Cys Gly Asn Glu Pro Val Ser Val Arg Cys 50 60

Phe Thr Asp Asn Leu Ile Leu Cys Gln Glu Cys Asp Trp Asp Val His 65 70 75 80

Gly Ser Cys Ser Val Ser Asp Ala His Val Arg Ser Ala Val Glu Gly 85 90 95

Phe Ser Gly Cys Pro Ser Ala Leu Glu Leu Ala Ala Leu Trp Gly Leu 100 105 110

Asp Leu Glu Gln Gly Arg Lys Asp Glu Glu Asn Gln Val Pro Met Met 115 120 125

Ala Met Met Met Asp Asn Phe Gly Met Gln Leu Asp Ser Trp Val Leu 130 135 140

<210> 30 <211> 433 <212> PRT

<213> Arabidopsis thaliana

# mbi19 Sequence Listing.ST25

Gly 145	Ser	Asn	Glu	Leu	Ile 150	Val	Pro	Ser	Авр	Thr 155	Thr	Phe	Lys	Lys	Arg 160
Gly	Ser	Сув	Gly	Ser 165	Ser	Сув	Gly	Arg	Tyr 170	Lys	Gln	Val	Leu	Сув 175	Lys
Gln	Leu	Glu	Glu 180	Leu	Leu	Lys	Ser	Gly 185	Val	Val	Gly	Gly	Asp 190	Gly	Asp
Asp	Gly	Asp 195	Arg	Asp	Arg	Asp	Сув 200	Asp	Arg	Glu	Gly	Ala 205	Сув	Asp	Gly
Asp	Gly 210	Asp	Gly	Glu	Ala	Gly 215	Glu	Gly	Leu	Met	Val 220	Pro	Glu	Met	Ser
225			-		230					235				Gly	240
				245					250					Pro 255	
-			260					265					270	Ser	
		275					280					285		Lys	
	290					295					300			Asn	
305				•	310					315				Asp	320
Tyr	Lys	Arg	Ser	Thr 325	Ser	Gly	Gln	Val	Gln 330	Pro	Thr	Lys	Ser	Glu 335	Ser
			340					345					350	Ser	
Ser	Asp	Leu 355		Phe	Thr		His 360		Ala	Gly	Thr	Ser 365		Lys	Thr
Thr	Arg 370	Leu	Val	Ala	Thr	Lys 375	Ala	Asp	Leu	Glu	Arg 380	Leu	Ala	Gln	Asn
Arg 385	Gly	Авр	Ala	Met	Gln 390	Arg	Tyr	Lys	Glu	Lys 395	Arg	ГЛВ	Thr	Arg	Arg 400
Tyr	Asp	Lys	Thr	Ile 405	Arg	Tyr	Glu	Ser	Arg 410	Lys	Ala	Arg	Ala	Asp 415	Thr
Arg	Leu	Arg	Val 420	Arg	Gly	Arg	Phe	Val 425	Lys	Ala	Ser	Glu	Ala 430	Pro	Tyr

Pro .

# mbi19 Sequence Listing.ST25

<213> Arabidopsis thaliana													
<220> <221> CDS <222> (208)(1200) <223> G1141													
<400> 31													
300000000 000300030 0000300030003000300	60 .20												
garrage of a control of a contr	.80												
coordinate and a grant control of the control of th	34												
Met Ala Val Tyr Glu Gln Thr Gly Thr 1 5													
gag cag ccg aag aaa agg aaa tct agg gct cga gca ggt ggt tta acg Glu Gln Pro Lys Lys Arg Lys Ser Arg Ala Arg Ala Gly Gly Leu Thr 10 15 20 25	82												
gtg gct gat agg cta aag aag tgg aaa gag tac aac gag att gtt gaa Val Ala Asp Arg Leu Lys Lys Trp Lys Glu Tyr Asn Glu Ile Val Glu 30 35 40	30												
gct tcg gct gtt aaa gaa gga gag aaa ccg aaa cgc aaa gtt cct gcg Ala Ser Ala Val Lys Glu Gly Glu Lys Pro Lys Arg Lys Val Pro Ala 45 50 55	78												
aaa ggg tcg aag aaa ggt tgt atg aag ggt aaa gga gg	26												
tct cac tgt agt ttt aga gga gtt aga caa agg att tgg ggt aaa tgg Ser His Cys Ser Phe Arg Gly Val Arg Gln Arg Ile Trp Gly Lys Trp 75 80 85	74												
gtt gca gag att cga gaa ccg aaa ata gga act aga ctt tgg ctt ggt Val Ala Glu Ile Arg Glu Pro Lys Ile Gly Thr Arg Leu Trp Leu Gly 90 95 100 105	22												
act ttt cct acc gcg gaa aaa gct gct tcc gct tat gat gaa gcg gct Thr Phe Pro Thr Ala Glu Lys Ala Ala Ser Ala Tyr Asp Glu Ala Ala 110 115 120	70												
acc gct atg tac ggt tca ttg gct cgt ctt aac ttc cct cag tct gtt Thr Ala Met Tyr Gly Ser Leu Ala Arg Leu Asn Phe Pro Gln Ser Val 125 130 135	18												
ggg tct gag ttt act agt acg tct agt caa tct gag gtg tgt acg gtt Gly Ser Glu Phe Thr Ser Thr Ser Ser Gln Ser Glu Val Cys Thr Val 140 145 150	66												
gaa aat aag gcg gtt gtt tgt ggt gat gtt tgt gtg aag cat gaa gat Glu Asn Lys Ala Val Val Cys Gly Asp Val Cys Val Lys His Glu Asp 155 160 165	14												
act gat tgt gaa tct aat cca ttt agt cag att tta gat gtt aga gaa Thr Asp Cys Glu Ser Asn Pro Phe Ser Gln Ile Leu Asp Val Arg Glu 170 175 180 185	62												
gag tot tgt gga acc agg ccg gac agt tgc acg gtt gga cat caa gat Glu Ser Cys Gly Thr Arg Pro Asp Ser Cys Thr Val Gly His Gln Asp 190 195 200	10												
atg aat tot tog otg aat tac gat ttg otg tta gag ttt gag cag cag Met Asn Ser Ser Leu Asn Tyr Asp Leu Leu Glu Phe Glu Gln Gln Page 42	58												

205		mbil9 Sequence 210	Listing.ST25 215	
tat tgg ggc caa Tyr Trp Gly Gln 220	gtt ttg cag Val Leu Gln	gag aaa gag aaa Glu Lys Glu Lys 225	ccg aag cag gaa gaa Pro Lys Gln Glu Glu 230	906
gag gag ata cag Glu Glu Ile Gln 235	caa cag caa Gln Gln Gln 240	cag gaa cag caa Gln Glu Gln Gln	cag caa cag ctg caa Gln Gln Gln Leu Gln 245	954
ccg gat ttg ctt Pro Asp Leu Leu 250	act gtt gca Thr Val Ala 255	gat tac ggt tgg Asp Tyr Gly Trp 260	cct tgg tct aat gat Pro Trp Ser Asn Asp 265	1002
att gta aat gat Ile Val Asn Asp	cag act tct Gln Thr Ser 270	tgg gat cct aat Trp Asp Pro Asn 275	gag tgc ttt gat att Glu Cys Phe Asp Ile 280	1050
aat gaa ctc ctt Asn Glu Leu Leu 285	gga gat ttg Gly Asp Leu	aat gaa cct ggt Asn Glu Pro Gly 290	ccc cat cag agc caa Pro His Gln Ser Gln 295	1098
gac caa aac cac Asp Gln Asn His 300	gta aat tct Val Asn Ser	ggt agt tat gat Gly Ser Tyr Asp 305	ttg cat ccg ctt cat Leu His Pro Leu His 310	1146
ctc gag cca cac Leu Glu Pro His 315	gat ggt cac Asp Gly His 320	gag ttc aat ggt Glu Phe Asn Gly	ttg agt tct ctg gat Leu Ser Ser Leu Asp 325	1194
att tga gagttctg Ile 330	gag gcaatggto	cc tacaagacta caa	acataatc tttggattga	1250
tcataggaga aacaa	agaaat aggtg	ttaat gatctgattc	acaatgaaaa aatatttaat	1310
aactctatag tttt	tgttct ttccti	tggat catgaactgt	tgcttctcat ctattgagtt	1370
aatatagcga atago	cagagt ttctc	tcata aaaaaaaaaa	aaa .	1413
<210> 32 <211> 330 <212> PRT <213> Arabidops	sis thaliana			
<400> 32				
Met Ala Val Tyr 1	Glu Gln Thr 5	Gly Thr Glu Gln 10	Pro Lys Lys Arg Lys 15	
Ser Arg Ala Arg 20	Ala Gly Gly	Leu Thr Val Ala 25	Asp Arg Leu Lys Lys 30	
Trp Lys Glu Tyr 35	Asn Glu Ile	Val Glu Ala Ser 40	Ala Val Lys Glu Gly 45	
Glu Lys Pro Lys 50	Arg Lys Val	Pro Ala Lys Gly	Ser Lys Lys Gly Cys 60	
Met Lys Gly Lys 65	Gly Gly Pro 70	Asp Asn Ser His 75	Cys Ser Phe Arg Gly 80	

Lys Ile Gly Thr Arg Leu Trp Leu Gly Thr Phe Pro Thr Ala Glu Lys 100  $$105\$ 

Val Arg Gln Arg Ile Trp Gly Lys Trp Val Ala Glu Ile Arg Glu Pro 95  $\phantom{000}$ 

mbil9 Sequence Listing.ST25

Ala Ala Ser Ala Tyr Asp Glu Ala Ala Thr Ala Met Tyr Gly Ser Leu

Ala Arg Leu Asn Phe Pro Gln Ser Val Gly Ser Glu Phe Thr Ser Thr

Ser Ser Gln Ser Glu Val Cys Thr Val Glu Asn Lys Ala Val Val Cys

Gly Asp Val Cys Val Lys His Glu Asp Thr Asp Cys Glu Ser Asn Pro

Phe Ser Gln Ile Leu Asp Val Arg Glu Glu Ser Cys Gly Thr Arg Pro

Asp Ser Cys Thr Val Gly His Gln Asp Met Asn Ser Ser Leu Asn Tyr

Asp Leu Leu Glu Phe Glu Gln Gln Tyr Trp Gly Gln Val Leu Gln

Glu Lys Glu Lys Pro Lys Gln Glu Glu Glu Glu Ile Gln Gln Gln 225 230 235 240

Gln Glu Gln Gln Gln Gln Leu Gln Pro Asp Leu Leu Thr Val Ala

Asp Tyr Gly Trp Pro Trp Ser Asn Asp Ile Val Asn Asp Gln Thr Ser 260 265 270

Trp Asp Pro Asn Glu Cys Phe Asp Ile Asn Glu Leu Leu Gly Asp Leu

Asn Glu Pro Gly Pro His Gln Ser Gln Asp Gln Asn His Val Asn Ser

Gly Ser Tyr Asp Leu His Pro Leu His Leu Glu Pro His Asp Gly His

Glu Phe Asn Gly Leu Ser Ser Leu Asp Ile

<210> 33 <211> 818 <212> DNA <213> Arabidopsis thaliana

<220>

<221> CDS <222> (18)..(641) <223> G46

ttctctcttt cgtaaaa atg gcg agt ttt gag gaa agc tct gat ttg gaa Met Ala Ser Phe Glu Glu Ser Ser Asp Leu Glu

gct ata cag agc cat ctc tta gaa gac ttg ttg gtt tgt gat ggt ttc Ala Ile Gln Ser His Leu Leu Glu Asp Leu Leu Val Cys Asp Gly Phe

50

98

							mbi	19 S	eque	nce	List	ing.	ST25			
			15					20	-			_	25			
								tct Ser								146
								gaa Glu								194
								ttg Leu								242
								tca Ser								290
								agg Arg 100								338
aga Arg	cat His	tac Tyr 110	cga Arg	gga Gly	gtg Val	agg Arg	cga Arg 115	agg Arg	ccg Pro	tgg Trp	999 Gly	aaa Lys 120	ttt Phe	gca Ala	gca Ala	386
gag Glu	att Ile 125	cga Arg	gat Asp	cca Pro	gca Ala	aag Lys 130	aaa Lys	gga Gly	tca Ser	aga Arg	atc Ile 135	tgg Trp	cta Leu	gga Gly	aca Thr	434
ttt Phe 140	gag Glu	agt Ser	gat Asp	gtt Val	gat Asp 145	gct Ala	gca Ala	aga Arg	gcc Ala	tat Tyr 150	gac Asp	tgt Cys	gca Ala	gct Ala	ttc Phe 155	482
aag Lys	ctc Leu	cgg Arg	gga Gly	aga Arg 160	aaa Lys	gcc Ala	gtg Val	ctc Leu	aac Asn 165	ttc Phe	cct Pro	ctt Leu	gac Asp	gcc Ala 170	Gly 999	530
aaa Lys	tat Tyr	gaa Glu	gct Ala 175	cca Pro	gcg Ala	aat Asn	tca Ser	gga Gly 180	agg Arg	aaa Lys	agg Arg	aag Lys	aga Arg 185	agt Ser	gat Asp	578
gtg Val	cat His	gaa Glu 190	gag Glu	ctt Leu	caa Gln	aga Arg	act Thr 195	cag Gln	agc Ser	aat Asn	tca Ser	tct Ser 200	tca Ser	tct Ser	tcc Ser	626
Cys				tag	cata	attaa	iga g	gtgtg	gagca	ag tt	tcct	taag	g ttg	gtata	aag	681
taat	tgta	aca g	gagga	aaac	ga at	tgtg	tage	ttt	agto	gtgc	ttg	aagt	tg (	caaca	aatgt	741
gtat	ggat	gt t	ctgt	ttct	t ca	atgto	ccta	a aga	ittta	agaa	acat	ctto	ett a	attto	ccaaga	801
aaaa	aaaa	aa a	aaaa	aaa									•			818
<210 <211 <212 <213	> 2 > E	34 207 PRT Arabi	dopa	sis t	hali	iana										
<400	> 3	34														
		_				_	_	_	_			-1.	<b>-3</b> .	_		

Leu Leu Glu Asp Leu Leu Val Cys Asp Gly Phe Met Gly Asp Phe Asp 20 25 Phe Met Gly Asp Phe Asp 30

Phe Asp Ala Ser Phe Val Ser Gly Leu Trp Cys Ile Glu Pro His Val 35

mbil9 Sequence Listing.ST25

							mbi	19 S	eque	nce	List	ing.	ST25			
Pro	Lys 50	Gln	Glu	Pro	Asp	Ser 55	Pro	Val	Leu	Asp	Pro 60	Asp	Ser	Phe	Val	
Asn 65	Glu	Phe	Leu	Gln	Val 70	Glu	Gly	Glu	Ser	Ser 75	Ser	Ser	Ser	Ser	Pro 80	
Glu	Leu	Asn	Ser	Ser 85	Ser	Ser	Thr	Tyr	Glu 90	Thr	Asp	Gln	Ser	Val 95	Lys	
Lys	Ala	Glu	Arg 100	Phe	Glu	Glu	Glu	Val 105	Asp	Ala	Arg	His	Tyr 110	Arg	Gly	
Val	Arg	Arg 115	Arg	Pro	Trp	Gly	Lys 120	Phe	Ala	Ala	Glu	Ile 125	Arg	Asp	Pro	
Ala	Lys 130	Lys	Gly	Ser	Arg	Ile 135	Trp	Leu	Gly	Thr	Phe 140	Glu	Ser	Asp	Val	
Asp 145	Ala	Ala	Arg	Ala	Tyr 150	Asp	Cys	Ala	Ala	Phe 155	Lys	Leu	Arg	Gly	Arg 160	
Lys	Ala	Val	Leu	Авп 165	Phe	Pro	Leu	Asp	Ala 170	Gly	ГÀа	Tyr	Glu	Ala 175	Pro	
Ala	Asn	Ser	Gly 180	Arg	Lys	Arg	Lys	Arg 185	Ser	Asp	Val	His	Glu 190	Glu	Leu	
Gln	Arg	Thr 195	Gln	Ser	Asn	Ser	Ser 200	Ser	Ser	Ser	Cys	Asp 205	Ala	Phe		
<210 <211 <212 <213	L> 1 2> 1	35 1195 ONA Arabi	dops	sis t	:hali	iana										
<220 <221 <222 <223	)> L> ( 2> +	DS	(98													
<400		35 aac d	caaaa	atcac	ct aa	agaq	qaqa	a aqa	attqo	ctaa	agtt	tgat	aa a	acat	tccaa	60
	a at	g go	et ga	at ag	g at	c aa	aa gg	gt co	a to	gg ag	gt co	et ga	aa ga	aa ga	ac gag sp Glu 15	110
							aaa Lys									158
att Ile	agc Ser	aaa Lys	tct Ser 35	att Ile	ccc Pro	ggt Gly	aga Arg	tcg Ser 40	999 Gly	aaa Lys	tcg Ser	tgt Cys	cgt Arg 45	tta Leu	cgg Arg	206
tgg Trp	tgc Сув	aac Asn 50	cag Gln	ctt Leu	tcg Ser	ccg Pro	caa Gln 55	gtt Val	gag Glu	cat His	cgg Arg	ccg Pro 60	ttt Phe	tcg Ser	gct Ala	254
							cgt Arg									302
•																

PCT/US00/31414 WO 01/35725

		mbile comic	nas Listina CT	25	
tgg gcg acg att Trp Ala Thr Ile 80	gct cgt ctt Ala Arg Leu 85	ctc aac ggt	nce Listing.ST cgt acg gac a Arg Thr Asp A 90	ac gcc gtg	350
aag aat cac tgo Lys Asn His Trp	aac tcg acg Asn Ser Thr 100	ctc aag agg Leu Lys Arg 105	aaa tgc ggc g Lys Cys Gly G	gt tac gac ly Tyr Asp 110	398
cat cgg ggt tac His Arg Gly Tyr 115	Asp Gly Ser	gag gat cat Glu Asp His 120	Arg Pro Val L	ag aga tcg . ys Arg Ser 25	446
gtg agt gcg gga Val Ser Ala Gly 130					494
gga agc cca act Gly Ser Pro Thr 145	gga tct gat Gly Ser Asp 150	gtc agt gat Val Ser Asp	tca agt act a Ser Ser Thr I 155	tc ccg ata le Pro Ile	542
tta cct tcc gtt Leu Pro Ser Val 160					590
gtg cta ccg ctt Val Leu Pro Leu					638
act tcg tta ago Thr Ser Leu Ser 195	Leu Ser Leu		Asp Val Ser G		686
aac cgt agc cac Asn Arg Ser His 210	gag tca acg Glu Ser Thr	aat atc aac Asn Ile Asn 215	aac acc act t Asn Thr Thr S 220	cg agc cgc er Ser Arg	734
cac aac cac aac His Asn His Asn 225	aat acg gtg Asn Thr Val 230	tcg ttt atg Ser Phe Met	ccg ttt agt g Pro Phe Ser G 235	gt ggg ttt ly Gly Phe	782
aga ggt gcg att Arg Gly Ala Ile 240	gag gaa atg Glu Glu Met 245	ggg aag tct Gly Lys Ser	ttt ccc ggt a Phe Pro Gly A 250	ac gga ggc sn Gly Gly 255	830
gag ttt atg gcg Glu Phe Met Ala					878
tac atg acg gag Tyr Met Thr Glu 275	Met Gln Arg		Gly Gly Phe V		926
ttc att gat aat Phe Ile Asp Asn 290	ggc atg att Gly Met Ile	ccg atg agt Pro Met Ser 295	caa att gga g Gln Ile Gly V 300	tt ggg aga al Gly Arg	974
atc gag tag aca Ile Glu 305	aagtgag attat	tagga aactgt	ttaa attggaga	ag	1023
aagaaaaatg ctct	gttttt ttctcc	tttg gattagg	ctt aagaatttt	g ggttttaagg	1083
aaatgtatag agga	aatcga gtgaac	aaag ctcgaga	gct ggggacgta	g tgacgaagac	1143
gaagatcaaa tttc	tcttaa gctatt	cagg aaaataa	aat aaatttta	t tt	1195

Met Ala Asp Arg Ile Lys Gly Pro Trp Ser Pro Glu Glu Asp Glu Gln

<210> 36 <211> 305 <212> PRT <213> Arabidopsis thaliana

<400> 36

mbil9 Sequence Listing.ST25

Leu Arg Arg Leu Val Val Lys Tyr Gly Pro Arg Asn Trp Thr Val Ile  $\frac{25}{20}$ 

Ser Lys Ser Ile Pro Gly Arg Ser Gly Lys Ser Cys Arg Leu Arg Trp 35 40 45

Cys Asn Gln Leu Ser Pro Gln Val Glu His Arg Pro Phe Ser Ala Glu 50 60

Glu Asp Glu Thr Ile Ala Arg Ala His Ala Gln Phe Gly Asn Lys Trp 65 70 75 80

Ala Thr Ile Ala Arg Leu Leu Asn Gly Arg Thr Asp Asn Ala Val Lys  $85 \hspace{1.5cm} 90 \hspace{1.5cm} 95$ 

Asn His Trp Asn Ser Thr Leu Lys Arg Lys Cys Gly Gly Tyr Asp His 100 105 110

Arg Gly Tyr Asp Gly Ser Glu Asp His Arg Pro Val Lys Arg Ser Val 115 120 125

Ser Ala Gly Ser Pro Pro Val Val Thr Gly Leu Tyr Met Ser Pro Gly 130 135 140

Ser Pro Thr Gly Ser Asp Val Ser Asp Ser Ser Thr Ile Pro Ile Leu 145 150 160

Pro Ser Val Glu Leu Phe Lys Pro Val Pro Arg Pro Gly Ala Val Val

Leu Pro Leu Pro Ile Glu Thr Ser Ser Phe Ser Asp Asp Pro Pro Thr 180 185 190

Ser Leu Ser Leu Pro Gly Ala Asp Val Ser Glu Glu Ser Asn 195 200 205

Arg Ser His Glu Ser Thr Asn Ile Asn Asn Thr Thr Ser Ser Arg His 210 215 220

Asn His Asn Asn Thr Val Ser Phe Met Pro Phe Ser Gly Gly Phe Arg 225 230 240

Gly Ala Ile Glu Glu Met Gly Lys Ser Phe Pro Gly Asn Gly Glu 245 250 255

Phe Met Ala Val Val Glu Glu Met Ile Lys Ala Glu Val Arg Ser Tyr 260 265 270

Met Thr Glu Met Gln Arg Asn Asn Gly Gly Gly Phe Val Gly Gly Phe 275 280 285

Ile Asp Asn Gly Met Ile Pro Met Ser Gln Ile Gly Val Gly Arg Ile 290 . 295 300

# mbi19 Sequence Listing.ST25

Glu 305														
<210> 37 <211> 989 <212> DNA <213> Arabidopsis thaliana														
<220> <221> CDS <222> (21)(983) <223> G227														
<pre>&lt;400&gt; 37 gtaccgtcga cgatccggcg atg tca aac ccg acc cgt aag aat atg gag agg</pre>														
att aaa ggt cca tgg agt cca gaa gaa gat gat ctg ttg cag agg ctt Ile Lys Gly Pro Trp Ser Pro Glu Glu Asp Asp Leu Leu Gln Arg Leu 15 20 25	101													
gtt cag aaa cat ggt ccg agg aac tgg tct ttg att agc aaa tca atc Val Gln Lys His Gly Pro Arg Asn Trp Ser Leu Ile Ser Lys Ser Ile 30 35 40	149													
cct gga cgt tcc ggc aaa tct tgt cgt ctc cgg tgg tgt aac cag cta Pro Gly Arg Ser Gly Lys Ser Cys Arg Leu Arg Trp Cys Asn Gln Leu 45 50 55	197													
tct ccg gag gta gag cac cgt gct ttt tcg cag gaa gaa gac gag acg Ser Pro Glu Val Glu His Arg Ala Phe Ser Gln Glu Glu Asp Glu Thr 60 70 75	245													
att att cga gct cac gct cgg ttt ggt aac aag tgg gct acg atc tct Ile Ile Arg Ala His Ala Arg Phe Gly Asn Lys Trp Ala Thr Ile Ser 80 85 90	293													
cgt ctt ctc aat gga cga acc gat aac gct atc aag aat cat tgg aac Arg Leu Leu Asn Gly Arg Thr Asp Asn Ala Ile Lys Asn His Trp Asn 95 100 105	341													
tcg acg ctg aag cga aaa tgc agc gtc gaa ggg caa agt tgt gat ttt Ser Thr Leu Lys Arg Lys Cys Ser Val Glu Gly Gln Ser Cys Asp Phe 110 115 120	389													
ggt ggt aat gga ggg tat gat ggt aat tta gga gaa ga	437													
aaa cgt acg gcg agt ggt ggt ggt gtc tcg act ggc ttg tat atg Lys Arg Thr Ala Ser Gly Gly Gly Val Ser Thr Gly Leu Tyr Met 140 145 150 155	485													
agt ccc gga agt cca tcg gga tct gac gtc agc gag caa tct agt ggt Ser Pro Gly Ser Pro Ser Gly Ser Asp Val Ser Glu Gln Ser Ser Gly 160 165 170	533													
ggt gca cac gtg ttt aaa cca acg gtt aga tct gag gtt aca gcg tca Gly Ala His Val Phe Lys Pro Thr Val Arg Ser Glu Val Thr Ala Ser 175 180 185	581													
tcg tct ggt gaa gat cct cca act tat ctt agt ttg tct ctt cct tgg Ser Ser Gly Glu Asp Pro Pro Thr Tyr Leu Ser Leu Ser Leu Pro Trp 190 195 200	629													
act gac gag acg gtt cga gtc aac gag ccg gtt caa ctt aac cag aat Thr Asp Glu Thr Val Arg Val Asn Glu Pro Val Gln Leu Asn Gln Asn 205 210 215	677													
acg gtt atg gac ggt ggt tat acg gcg gag ctg ttt ccg gtt aga aag Thr Val Met Asp Gly Gly Tyr Thr Ala Glu Leu Phe Pro Val Arg Lys 220 235	725													
Page 49														

	_		
mbi19	Semience	Listing	. ST25

								mbi	19 S	eque	nce	List	ing.	ST25			
ga G1	a g u G	gag Slu	caa Gln	gtg Val	gaa Glu 240	gta Val `	gaa Glu	gaa Glu	gaa Glu	gaa Glu 245	gcg Ala	aag Lys	999 Gly	ata Ile	tct Ser 250	ggt Gly	773
						ttc Phe											821
ga G1	g g u V	gtg /al	agg Arg 270	agt Ser	tac Tyr	atg Met	gcg Ala	gat Asp 275	tta Leu	cag Gln	cga Arg	gga Gly	aac Asn 280	gtc Val	ggt Gly	ggt Gly	869
ag Se	r S	gt Ser 185	tct Ser	ggc Gly	ggc Gly	gga Gly	ggt Gly 290	ggc Gly	ggt Gly	tcg Ser	tgt Cys	atg Met 295	cca Pro	caa Gln	agt Ser	gta Val	917
	n S					999 Gly 305											965
				atg Met		tag	gcgg	gcc									989
<2 <2	10> 11> 12> 13>	. 3	8 20 RT Tabi	idops	sis 1	thali	iana										
<4	00>	3	8														
Me 1	t S	er	Asn	Pro	Thr 5	Arg	Lys	Asn	Met	Glu 10	Arg	Ile	Lys	Gly	Pro 15	Trp	
Se	r P	ro	Glu	Glu 20	Asp	Asp	Leu	Leu	Gln 25	Arg	Leu	Val	Gln	Lys 30	His	Gly	
Pr	o A	rg	Asn 35	Trp	Ser	Leu	Ile	Ser 40	Lys	Ser	Ile	Pro	Gly 45	Arg	Ser	Gly	
Ly		er 0	Сув	Arg	Leu	Arg	Trp 55	Сув	Asn	Gln	Leu	Ser 60	Pro	Glu	Val	Glu	
Ні 65		rg	Ala	Phe	Ser	Gln 70	Glu	Glu	Авр	Glu	Thr 75	Ile	Ile	Arg	Ala	His 80	
Al	a A	rg	Phe	Gly	Asn 85	Lys	Trp	Ala	Thr	Ile 90	Ser	Arg	Leu	Leu	Asn 95	Gly	
Ar	g T	hr	Asp	Asn 100	Ala	Ile	Lys	Asn	His 105	Trp	Asn	Ser	Thr	Leu 110	Lys	Arg	
Ly	вC		Ser 115	Val	Glu	Gly	Gln	Ser 120	Сув	Asp	Phe	Gly	Gly 125	Asn	Gly	Gly	
ту		.sp 30	Gly	Asn	Leu	Gly	Glu 135	Glu	Gln	Pro	Leu	Lys 140	Arg	Thr	Ala	Ser	
G1 14		ly	Gly	Gly	Val	Ser 150	Thr	Gly	Leu	Tyr	Met 155	Ser	Pro	Gly	Ser	Pro 160	

Ser Gly Ser Asp Val Ser Glu Gln Ser Ser Gly Gly Ala His Val Phe 165  $\,$  170  $\,$  175  $\,$ 

# mbil9 Sequence Listing.ST25

Lys Pro Thr Val Arg Ser Glu Val Thr Ala Ser Ser Ser Gly Glu Asp 180 185 190						
Pro Pro Thr Tyr Leu Ser Leu Ser Leu Pro Trp Thr Asp Glu Thr Val 195 200 205						
Arg Val Asn Glu Pro Val Gln Leu Asn Gln Asn Thr Val Met Asp Gly 210 215 220						
Gly Tyr Thr Ala Glu Leu Phe Pro Val Arg Lys Glu Glu Gln Val Glu 225 230 240						
Val Glu Glu Glu Ala Lys Gly Ile Ser Gly Gly Phe Gly Glu 245 250 255						
Phe Met Thr Val Val Gln Glu Met Ile Arg Thr Glu Val Arg Ser Tyr 260 265 270						
Met Ala Asp Leu Gln Arg Gly Asn Val Gly Gly Ser Ser Ser Gly Gly 275 280 285						
Gly Gly Gly Ser Cys Met Pro Gln Ser Val Asn Ser Arg Arg Val 290 295 300						
Gly Phe Arg Glu Phe Ile Val Asn Gln Ile Gly Ile Gly Lys Met Glu 305 310 315 320						
<210> 39 <211> 994 <212> DNA <213> Arabidopsis thaliana						
<220> <221> CDS <222> (140)(889) <223> G1307						
<400> 39 cccttattgg gcntnancgn ccncccggca ggtctagnnn tnancgcccg cgtccttctn 60						
ccattttacn cncttgcngc ccacccttgt atntcntttt ntnngtgntn tttttcntga 120						
gggggcaacg gaaaaaaga atg gga aga gca cca tgt tgt gag aaa atg ggg 172 Met Gly Arg Ala Pro Cys Cys Glu Lys Met Gly 1 5 10						
gtg aag aga gga cca tgg act cct gaa gaa gat caa atc ttg atc aat Val Lys Arg Gly Pro Trp Thr Pro Glu Glu Asp Gln Ile Leu Ile Asn 15 20 25	220					
tat att cat ctt tat ggt cat tct aat tgg cga gct ctc cca aaa cac Tyr Ile His Leu Tyr Gly His Ser Asn Trp Arg Ala Leu Pro Lys His 30 35 40	268					
gca ggt tta ctt aga tgt ggg aaa agt tgc aga ctt ggt tgg atc aat Ala Gly Leu Leu Arg Cys Gly Lys Ser Cys Arg Leu Gly Trp Ile Asn 45 50 55	316					
tat ctt aga cca gac att aaa cgt ggc aat ttc act cct caa gaa gaa Tyr Leu Arg Pro Asp Ile Lys Arg Gly Asn Phe Thr Pro Gln Glu Glu 60 65 70 75	364					
caa act att atc aat ctg cat gaa agc tta ggc aac aga tgg tct gcg Page 51	412					

Gln Thr Ile I	le Asn Leu His 80	mbi19 Sequence Glu Ser Leu Gly 85				
Ile Ala Ala L		cga acc gac aat Arg Thr Asp Asn 100				
		aga ctc agc aaa Arg Leu Ser Lys 115				
gga gac acc a Gly Asp Thr L 125	aa gac gtt aac ys Asp Val Asn 130	gga att aac gag Gly Ile Asn Glu	acc aca aat ga Thr Thr Asn Gl 135	a gac 556 u Asp		
aaa gga tct g Lys Gly Ser V 140	tg ata gtc gac al Ile Val Asp 145	aca gcc tct tta Thr Ala Ser Leu 150	caa caa ttt to Gln Gln Phe Se	t aat 604 r Asn 155		
agt att aca a Ser Ile Thr T	ca ttt gat att hr Phe Asp Ile 160	tca aat gat aac Ser Asn Asp Asn 165	aag gac gat at Lys Asp Asp Il 17	e Met		
Ser Tyr Glu A	at att tct gcc sp Ile Ser Ala 75	ttg ata gat gat Leu Ile Asp Asp 180	agt ttt tgg tc Ser Phe Trp Se 185	g gac 700 r Asp		
gtc ata tcg g Val Ile Ser V 190	ta gat aat tcg al Asp Asn Ser	aat aag aat gag Asn Lys Asn Glu 195	aag aag ata ga Lys Lys Ile Gl 200	g gat 748 u Asp		
tgg gaa gga t Trp Glu Gly L 205	tg atc gat aga eu Ile Asp Arg 210	aat agt aaa aaa Asn Ser Lys Lys	tgt agc tat ag Cys Ser Tyr Se 215	t aat 796 r Asn		
tct aag ttg t Ser Lys Leu T 220	at aat gat gac yr Asn Asp Asp 225	atg gag ttt tgg Met Glu Phe Trp 230	ttt gat gtt tt Phe Asp Val Ph	c act 844 e Thr 235		
		ttt tcc gac ata Phe Ser Asp Ile 245		a 889		
ttttgatttt gattttgtgt tgtttttgtc gttaagactt tgaaagtctt tttgtaatcc 949						
				994		
aaatgaataa attootttto tttttaaaaa aaaaaaaaaa						
<210> 40 <211> 249 <212> PRT <213> Arabidopsis thaliana						
<400> 40						
Met Gly Arg A 1	la Pro Cys Cys 5	Glu Lys Met Gly 10	Val Lys Arg Gl 15			
Trp Thr Pro G		Ile Leu Ile Asn 25	Tyr Ile His Le 30	u Tyr		
Gly His Ser A 35	sn Trp Arg Ala	Leu Pro Lys His 40	Ala Gly Leu Le 45	u Arg		
Cys Gly Lys S	er Cys Arg Leu 55	Gly Trp Ile Asn	Tyr Leu Arg Pr	o Asp		

mbil9 Sequence Listing.ST25 Leu His Glu Ser Leu Gly Asn Arg Trp Ser Ala Ile Ala Ala Lys Leu Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn Val Trp His Thr His Leu Lys Lys Arg Leu Ser Lys Asn Leu Asn Asn Gly Gly Asp Thr Lys Asp 115 120 125 Val Asn Gly Ile Asn Glu Thr Thr Asn Glu Asp Lys Gly Ser Val Ile Val Asp Thr Ala Ser Leu Gln Gln Phe Ser Asn Ser Ile Thr Thr Phe Asp Ile Ser Asn Asp Asn Lys Asp Asp Ile Met Ser Tyr Glu Asp Ile Ser Ala Leu Ile Asp Asp Ser Phe Trp Ser Asp Val Ile Ser Val Asp 180 185 190 Asn Ser Asn Lys Asn Glu Lys Lys Ile Glu Asp Trp Glu Gly Leu Ile Asp Arg Asn Ser Lys Lys Cys Ser Tyr Ser Asn Ser Lys Leu Tyr Asn 210 215 220 Asp Asp Met Glu Phe Trp Phe Asp Val Phe Thr Ser Asn Arg Arg Ile 225 230 235 Glu Glu Phe Ser Asp Ile Pro Glu Phe 245 <210> 41 <211> 891 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (1)..(891) <223> G1327 atg ggg aaa gga aga gca cca tgc tgc gac aag aac aaa gtg aag aga Met Gly Lys Gly Arg Ala Pro Cys Cys Asp Lys Asn Lys Val Lys Arg ggg cca tgg agc cct caa gaa gat ctc act ctc act ttt att caa Gly Pro Trp Ser Pro Gln Glu Asp Leu Thr Leu Ile Thr Phe Ile Gln 96 aaa cat ggc cat caa aac tgg aga tct ctt ccc aag ctt gct gga ttg Lys His Gly His Gln Asn Trp Arg Ser Leu Pro Lys Leu Ala Gly Leu ttg aga tgt ggg aaa agt tgc cga cta aga tgg ata aac tat ctg aga Leu Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp Ile Asn Tyr Leu Arg 50 55 60192 ccg gac gtg aag cga ggc aac ttt agc aaa aag gag gaa gat gct atc Pro Asp Val Lys Arg Gly Asn Phe Ser Lys Lys Glu Glu Asp Ala Ile 65 70 75 80 240

mbi19 Sequence Listing.ST25

									-4							
att Ile	cac His	tac Tyr	cat His	caa Gln 85	acc Thr	ctt Leu	gga Gly	aac Asn	aag Lys 90	tgg Trp	tca Ser	aag Lys	atc Ile	gcg Ala 95	tcc Ser	288
ttc Phe	ttg Leu	ccg Pro	gga Gly 100	aga Arg	act Thr	gac Asp	aac Asn	gag Glu 105	atc Ile	aaa Lys	aac Asn	gtg Val	tgg Trp 110	aac Asn	acg Thr	336
cat His	ctc Leu	aag Lys 115	aaa Lys	cga Arg	ctc Leu	act Thr	cca Pro 120	tct Ser	tct Ser	tct Ser	tct Ser	tca Ser 125	tcc Ser	ctc Leu	tct Ser	384
agc Ser	act Thr 130	cat His	gac Asp	caa Gln	agc Ser	aca Thr 135	aaa Lys	gca Ala	gat Asp	cat His	дас Авр 140	aag Lys	aac Asn	tgt Cys	gac Asp	432
999 Gly 145	gct Ala	caa Gln	gaa Glu	gaa Glu	ata Ile 150	cat His	tca Ser	999 Gly	tta Leu	aat Asn 155	gag Glu	agc Ser	caa Gln	aac Asn	tca Ser 160	480
gct Ala	act Thr	tcg Ser	tca Ser	cat His 165	cac His	caa Gln	ggc Gly	gag Glu	tgt Cys 170	atg Met	cac His	aca Thr	aaa Lys	cca Pro 175	gag Glu	528
ctt Leu	cat His	gag Glu	gtt Val 180	aat Asn	gga Gly	ctc Leu	aac Asn	gag Glu 185	atc Ile	cag Gln	ttc Phe	ctg Leu	ctc Leu 190	gac Asp	cat His	576
gat Asp	gac Asp	ttt Phe 195	gat Asp	gat Asp	ata Ile	acc Thr	tct Ser 200	gag Glu	ttt Phe	ctt Leu	cag Gln	gat Asp 205	aac Asn	gat Asp	atc Ile	624
tta Leu	ttt Phe 210	ccg Pro	cta Leu	gac Asp	tct Ser	ctt Leu 215	ctt Leu	cat His	aac Asn	cac His	caa Gln 220	act Thr	cac His	att Ile	tca Ser	672
acc Thr 225	caa Gln	gaa Glu	atg Met	act Thr	cga Arg 230	gag Glu	gta Val	acc Thr	aaa Lys	tcg Ser 235	caa Gln	tca Ser	ttt Phe	gat Asp	cat His 240	720 .
cct Pro	caa Gln	ccg Pro	gat Asp	atc Ile 245	cca Pro	tgc Cys	gga Gly	ttt Phe	gaa Glu 250	gac Asp	aca Thr	aac Asn	gaa Glu	gaa Glu 255	tcc Ser	768
gac Asp	ttg Leu	agg Arg	aga Arg 260	cag Gln	ctg Leu	gtt Val	gaa Glu	tca Ser 265	acc Thr	aca Thr	cct Pro	aac Asn	aat Asn 270	gag Glu	tac Tyr	816
gac Asp	gag Glu	tgg Trp 275	ttc Phe	aac Asn	ttc Phe	att Ile	gac Asp 280	aac Asn	caa Gln	act Thr	tac Tyr	ttt Phe 285	gat Asp	gat Asp	ttt Phe	864
aat Asn	ttc Phe 290	gtc Val	gga Gly	gaa Glu	gta Val	tgt Cys 295	cta Leu	tga								891
<210 <211 <212 <213	> 2  > 1	12 296 PRT Arabi	dops	sis t	hali	iana										
<400	)> 4	2														
Met	Gly	Lys	Gly	Arg	Ala	Pro	Сув	Cys	Asp	Lys	Asn	Lys	Val	Lув 15	Arg	

Met Gly Lys Gly Arg Ala Pro Cys Cys Asp Lys Asn Lys Val Lys Arg 1 5 10 15

Gly Pro Trp Ser Pro Gln Glu Asp Leu Thr Leu Ile Thr Phe Ile Gln 20 25 30

Lys His Gly His Gln Asn Trp Arg Ser Leu Pro Lys Leu Ala Gly Leu 35

PCT/US00/31414 WO 01/35725

### mbi19 Sequence Listing.ST25

Leu Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp Ile Asn Tyr Leu Arg 50 60

Pro Asp Val Lys Arg Gly Asn Phe Ser Lys Lys Glu Glu Asp Ala Ile 65 70 75 80

Ile His Tyr His Gln Thr Leu Gly Asn Lys Trp Ser Lys Ile Ala Ser 85 90 95

Phe Leu Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn Val Trp Asn Thr

His Leu Lys Lys Arg Leu Thr Pro Ser Ser Ser Ser Ser Ser Leu Ser

Ser Thr His Asp Gln Ser Thr Lys Ala Asp His Asp Lys Asn Cys Asp 130 135 140

Gly Ala Gln Glu Glu Ile His Ser Gly Leu Asn Glu Ser Gln Asn Ser

Ala Thr Ser Ser His His Gln Gly Glu Cys Met His Thr Lys Pro Glu

Leu His Glu Val Asn Gly Leu Asn Glu Ile Gln Phe Leu Leu Asp His 180 185 190

Asp Asp Phe Asp Asp Ile Thr Ser Glu Phe Leu Gln Asp Asn Asp Ile

Leu Phe Pro Leu Asp Ser Leu Leu His Asn His Gln Thr His Ile Ser 210 215 220

Thr Gln Glu Met Thr Arg Glu Val Thr Lys Ser Gln Ser Phe Asp His

Pro Gln Pro Asp Ile Pro Cys Gly Phe Glu Asp Thr Asn Glu Glu Ser 245 250 255

Asp Leu Arg Arg Gln Leu Val Glu Ser Thr Thr Pro Asn Asn Glu Tyr

Asp Glu Trp Phe Asn Phe Ile Asp Asn Gln Thr Tyr Phe Asp Asp Phe 275 280 285

Asn Phe Val Gly Glu Val Cys Leu 290 295

<210> 43 <211> 1237 <212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS <222> (73)..(954) <223> G673

## mbi19 Sequence Listing.ST25

<40 tct		43 cta	accc	cttc	tc t	cttc	agtc	t ct	ctct	ctct	aga	cgat	ctc	tatc	ttgaat	60
aaa	atac	cga	ta a M	tg a et T	cc to hr So	ca a er T	cc a hr A 5	at co sn P:	cg g ro V	tg g al V	tc g	cc g la G 1	lu V	ta a al I	ta ccg le Pro	111
gcg Ala	gaa Glu 15	act Thr	tct Ser	aca Thr	gat Asp	gct Ala 20	aca Thr	gag Glu	acg Thr	acg Thr	att Ile 25	gca Ala	acg Thr	acg Thr	gaa Glu	159
gct Ala 30	ggt Gly	gaa Glu	gca Ala	ccg Pro	gag Glu 35	aag Lys	aag Lys	gtg Val	agg Arg	aaa Lys 40	gct Ala	tat Tyr	aca Thr	atc Ile	acc Thr 45	207
aag Lys	tct Ser	aga Arg	gag Glu	agt Ser 50	tgg Trp	act Thr	gaa Glu	gga Gly	gaa Glu 55	cac His	gac Asp	aag Lys	ttt Phe	ctg Leu 60	gaa Glu	255
gct Ala	ctt Leu	caa Gln	ttg Leu 65	ttt Phe	gat Asp	cgt Arg	gac Asp	tgg Trp 70	aaa Lys	aag Lys	ata Ile	gaa Glu	gat Asp 75	ttt Phe	gtt Val	303
ggt Gly	tca Ser	aag Lys 80	aca Thr	gtt Val	att Ile	cag Gln	atc Ile 85	agg Arg	agc Ser	cat His	gcc Ala	caa Gln 90	aaa Lys	tac Tyr	ttt Phe	351
cta Leu	aag Lys 95	gtc Val	caa Gln	aaa Lys	aat Asn	999 Gly 100	act Thr	tta Leu	gca Ala	cat His	gtt Val 105	cca Pro	ccc Pro	cct Pro	agg Arg	399
cct Pro 110	Lys	cgc Arg	aaa Lys	gct Ala	gct Ala 115	cat His	cca Pro	tat Tyr	cct Pro	caa Gln 120	aag Lys	gca Ala	tcg Ser	aaa Lys	aat Asn 125	447
gct Ala	caa Gln	atg Met	tcg Ser	ctt Leu 130	cac His	gtt Val	tcc Ser	atg Met	tcc Ser 135	ttt Phe	cct Pro	act Thr	caa Gln	ata Ile 140	aat Asn	495
aac Asn	ctg Leu	cct Pro	gga Gly 145	tat Tyr	act Thr	cca Pro	tgg Trp	gat Asp 150	gat Asp	gat Asp	aca Thr	tct Ser	gca Ala 155	ttg Leu	tta Leu	543
aac Asn	att Ile	gct Ala 160	gta Val	agt Ser	G1 y 999	gtt Val	att Ile 165	cca Pro	cca Pro	gaa Glu	gat Asp	gaa Glu 170	ctt Leu	gat Asp	act Thr	591
ctt Leu	tgt Cys 175	gga Gly	gca Ala	gaa Glu	gtt Val	gat Asp 180	gtt Val	gga Gly	tca Ser	aat Asn	gac Asp 185	atg Met	ata Ile	agt Ser	gaa Glu	639
					tct Ser 195											687
gat Asp	tct Ser	aag Lys	ggt Gly	ttg Leu 210	aga Arg	ctg Leu	gcg Ala	aaa Lys	caa Gln 215	gct Ala	ccc Pro	tca Ser	atg Met	cat His 220	ggt Gly	735
					gag Glu											783
					cgc Arg											831
					ttg Leu											879
					gaa Glu				Ğlu		Val					927

270		275	шb	i19 S		e List	ing.S	ST25		285	
gaa ggt d Glu Gly H					ctgtt	tgtgc a	actca	acaa	ıg		974
ttatatato	ct totto		tcttgcto	g caa	caact	ct ctac	cago	ta t	caaa	itgcat	1034
cgtacggtt											1094
tgcagttt											1154
tggttacca	aa ttgta	tcttc t	ttttgata	a tta	tttt	tc atgo	cttt	tg t	aata	itgtaa	1214
gtttcttta	aa aaaaa	aaaaa a	aa								1237
<212> PF	93 RT	sis thal:	iana								
<400> 44	4										
Met Thr S	Ser Thr	Asn Pro 5	Val Val	Ala	Glu V 10	al Ile	Pro	Ala	Glu 15	Thr	
Ser Thr A	Asp Ala 20	Thr Glu	Thr Th	Ile 25	Ala T	hr Thr	Glu	Ala 30	Gly	Glu	
Ala Pro (	Glu Lys 35	Lys Val	Arg Lys	a Ala	Tyr T	hr Ile	Thr 45	Lys	Ser	Arg	
Glu Ser 1		Glu Gly	Glu His	Asp	Lys P	he Leu 60	Glu	Ala	Leu	Gln	
Leu Phe A	Asp Arg	Asp Trp 70	Lys Lys	Ile	Glu A	sp Phe 5	Val	Gly	Ser	Fys	
Thr Val	Ile Gln	Ile Arg 85	Ser His	a Ala	Gln L 90	ys Tyr	Phe	Leu	Lys 95	Val	
Gln Lys A	Asn Gly 100	Thr Leu	Ala Hi	Val 105	Pro P	ro Pro	Arg	Pro 110	Lys	Arg	
Lys Ala i	Ala His 115	Pro Tyr	Pro Gli 120	Lys	Ala S	er Lys	Asn 125	Ala	Gln	Met	
Ser Leu 1	His Val	Ser Met	Ser Pho	Pro	Thr G	ln Ile 140	Asn	Asn	Leu	Pro	
Gly Tyr 3	Thr Pro	Trp Asp 150		Thr		la Leu 55	Leu	Asn	Ile	Ala 160	
Val Ser (	Gly Val	Ile Pro 165	Pro Gl	a Asp	Glu L 170	eu Asp	Thr	Leu	Cys 175	Gly	
Ala Glu '	Val Asp 180	Val Gly	Ser As	185	Met I	le Ser	Glu	Thr 190	Ser	Pro	
Ser Ala	Ser Gly 195	Ile Gly	Ser Se		Arg T	hr Leu	Ser 205	Asp	Ser	Lys	

### mbi19 Sequence Listing.ST25

Gly Leu Arg Leu Ala Lys Gln Ala Pro Ser Met His Gly Leu Pro Asp 210

Phe Ala Glu Val Tyr Asn Phe Ile Gly Ser Val Phe Asp Pro Asp Ser 220

225 230 235 240

Lys Gly Arg Met Lys Lys Leu Lys Glu Met Asp Pro Ile Asn Phe Glu 245 250 255

Thr Val Leu Leu Met Arg Asn Leu Thr Val Asn Leu Ser Asn Pro 260 265 270

Asp Phe Glu Pro Thr Ser Glu Tyr Val Asp Ala Ala Glu Glu Gly His 275 280 285

Glu His Leu Ser Ser 290

<210> 45 <211> 1764 <212> DNA <213> Arabidopsis thaliana

<221> CDS <222> (1)..(1764) <223> G307

<223> G307

act tct tct tct tca tca tca atc tct aaa gat aag atg atg gtg
Thr Ser Ser Ser Ser Ser Ile Ser Lys Asp Lys Met Met Wal
20 25 30

aaa aaa gaa gaa gac ggt gga ggt aac atg gac gac gac ctt ctc gct Lys Lys Glu Glu Asp Gly Gly Asn Met Asp Asp Glu Leu Leu Ala 35 40 45

gtt tta ggt tac aaa gtt agg tca tcg gag atg gcg gag gtt gct ttg
Val Leu Gly Tyr Lys Val Arg Ser Ser Glu Met Ala Glu Val Ala Leu
50 55 60

aaa ctc gaa caa tta gag acg atg atg agt aat gtt caa gaa gat ggt
Lys Leu Glu Gln Leu Glu Thr Met Met Ser Asn Val Gln Glu Asp Gly
65 75 80

tta tct cat ctc gcg acg gat act gtt cat tat aat ccg tcg gag ctt
Leu Ser His Leu Ala Thr Asp Thr Val His Tyr Asn Pro Ser Glu Leu
85 90 95

tat tct tgg ctt gat aat atg ctc tct gag ctt aat cct cct ctt Tyr Ser Trp Leu Asp Asn Met Leu Ser Glu Leu Asn Pro Pro Pro Leu 100 105 110

ccg gcg agt tct aac ggt tta gat ccg gtt ctt cct tcg ccg gag att
Pro Ala Ser Ser Asn Gly Leu Asp Pro Val Leu Pro Ser Pro Glu Ile

tgt ggt ttt ccg gct tcg gat tat gac ctt aaa gtc att ccc gga aac
Cys Gly Phe Pro Ala Ser Asp Tyr Asp Leu Lys Val Ile Pro Gly Asn
130
140

gcg att tat cag ttt ccg gcg att gat tct tcg tct tcg tcg aat aat
Ala Ile Tyr Gln Phe Pro Ala Ile Asp Ser Ser Ser Ser Ser Asn Asn

Page 58

							mbi	19 S	eque	nce	List	ing.	ST25			
145					150				-	155					160	
cag Gln	aac Asn	aag Lys	cgt Arg	ttg Leu 165	aaa Lys	tca Ser	tgc Cys	tcg Ser	agt Ser 170	cct Pro	gat Asp	tct Ser	atg Met	gtt Val 175	aca Thr	528
tcg Ser	act Thr	tcg Ser	acg Thr 180	ggt Gly	acg Thr	cag Gln	att Ile	ggt Gly 185	gga Gly	gtc Val	ata Ile	gga Gly	acg Thr 190	acg Thr	gtg Val	576
acg Thr	aca Thr	acc Thr 195	acc Thr	acg Thr	aca Thr	acg Thr	acg Thr 200	gcg Ala	gcg Ala	gct Ala	gag Glu	tca Ser 205	act Thr	cgt Arg	tct Ser	624
gtt Val	atc Ile 210	ctg Leu	gtt Val	gac Asp	tcg Ser	caa Gln 215	gag Glu	aac Asn	ggt Gly	gtt Val	cgt Arg 220	tta Leu	gtc Val	cac His	gcg Ala	672
ctt Leu 225	atg Met	gct Ala	tgt Cys	gca Ala	gaa Glu 230	gca Ala	atc Ile	cag Gln	cag Gln	aac Asn 235	aat Asn	ttg Leu	act Thr	cta Leu	gcg Ala 240	720
gaa Glu	gct Ala	ctt Leu	gtg Val	aag Lys 245	caa Gln	atc Ile	gga Gly	tgc Cys	tta Leu 250	gct Ala	gtg Val	tct Ser	caa Gln	gcc Ala 255	gga Gly	768
gct Ala	atg Met	aga Arg	aaa Lys 260	gtg Val	gct Ala	act Thr	tac Tyr	ttc Phe 265	gcc Ala	gaa Glu	gct Ala	tta Leu	gct Ala 270	cgg Arg	cgg Arg	816
atc Ile	tac Tyr	cgt Arg 275	ctc Leu	tct Ser	ccg Pro	ccg Pro	cag Gln 280	aat Asn	cag Gln	atc Ile	gat Asp	cat His 285	tgt Cys	ctc Leu	tcc Ser	864
gat Asp	act Thr 290	ctt Leu	cag Gln	atg Met	cac His	ttt Phe 295	tac Tyr	gag Glu	act Thr	tgt Cys	cct Pro 300	tat Tyr	ctt Leu	aaa Lys	ttc Phe	912
gct Ala 305	cac His	ttc Phe	acg Thr	gcg Ala	aac Asn 310	caa Gln	gcg Ala	att Ile	ctc Leu	gaa Glu 315	gct Ala	ttt Phe	gaa Glu	ggt Gly	aag Lys 320	960
aag Lys	aga Arg	gta Val	cac Kis	gtc Val 325	att Ile	gat Asp	ttc Phe	tcg Ser	atg Met 330	aac Asn	caa Gln	ggt Gly	ctt Leu	caa Gln 335	tgg Trp	1008
cct Pro	gcg Ala	ctt Leu	atg Met 340	caa Gln	gct Ala	ctt Leu	gcg Ala	ctt Leu 345	cga Arg	gaa Glu	gga Gly	ggt Gly	cct Pro 350	cca Pro	act Thr	1056
ttc Phe	cgg Arg	tta Leu 355	acc Thr	gga Gly	att Ile	ggt Gly	cca Pro 360	Pro	gcg Ala	ccg Pro	gat Asp	aat Asn 365	Ser	gat Asp	cat His	1104
ctt Leu	cat His 370	gaa Glu	gtt Val	ggt Gly	tgt Cys	aaa Lys 375	tta Leu	gct Ala	cag Gln	ctt Leu	gcg Ala 380	gag Glu	gcg Ala	att Ile	cac His	1152
gta Val 385	gaa Glu	ttc Phe	gaa Glu	tac Tyr	cgt Arg 390	gga Gly	ttc Phe	gtt Val	gct Ala	aac Asn 395	agc Ser	tta Leu	gcc Ala	gat Asp	ctc Leu 400	1200
gat Asp	gct Ala	tcg Ser	atg Met	ctt Leu 405	gag Glu	ctt Leu	aga Arg	ccg Pro	agc Ser 410	gat Asp	acg Thr	gaa Glu	gct Ala	gtt Val 415	gcg Ala	1248
gtg Val	aac Asn	tct Ser	gtt Val 420	ttt Phe	gag Glu	cta Leu	cat His	aag Lys 425	ctc Leu	tta Leu	ggt Gly	cgt Arg	ccc Pro 430	ggt Gly	999 Gly	1296
ata Ile	gag Glu	aaa Lys 435	gtt Val	ctc Leu	ggc Gly	gtt Val	gtg Val 440	aaa Lys	cag Gln	att Ile	aaa Lys	ccg Pro 445	gtg Val	att Ile	ttc Phe	1344
acg	gtg	gtt	gag	caa	gaa	tcg	aac	cat		gga age		gtt	ttc	tta	gac	1392

Thr Val Va 450	l Glu Gl	n Glu	Ser 455							ST25 Phe	Leu	Asp	
cgg ttt ac Arg Phe Th 465	gaa to Glu Se	g tta r Leu 470	cat His	tat Tyr	tat Tyr	tcg Ser	act Thr 475	ctg Leu	ttt Phe	gat Asp	tcg Ser	ttg Leu 480	1440
gaa gga gt Glu Gly Va	ccg aa l Pro As 48	n Ser	caa Gln	gac Asp	aaa Lys	gtc Val 490	atg Met	tct Ser	gaa Glu	gtt Val	tac Tyr 495	tta Leu	1488
ggg aaa ca Gly Lys Gl													1536
gag aga ca Glu Arg Hi 51	3 Glu Th	g ttg r Leu	agt Ser	caa Gln 520	tgg Trp	gga Gly	aac Asn	cgg Arg	ttt Phe 525	ggt Gly	tcg Ser	tcc Ser	1584
ggt tta gc Gly Leu Ala 530	g ccg gc a Pro Al	a cat a His	ctt Leu 535	G1 y 999	tct Ser	aac Asn	gcg Ala	ttt Phe 540	aag Lys	caa Gln	gcg Ala	agt Ser	1632
atg ctt tte Met Leu Les 545													1680
agt aat gg	tgt tt Cys Le 56	u Met	ttg Leu	ggt Gly	tgg Trp	cac His 570	act Thr	cgc Arg	cca Pro	ctc Leu	att Ile 575	acc Thr	1728
acc tcc gc Thr Ser Ala								tga					1764
<210> 46 <211> 587 <212> PRT <213> Aral	oidopsis	thali	iana										
<211> 587 <212> PRT	oidopsis	thali	iana										
<211> 587 <212> PRT <213> Aral				Phe	Gln	Gly 10	Arg	Leu	Ser	Asn	His 15	Gly	
<211> 587 <212> PRT <213> Aral <400> 46 Met Lys Ara	Asp Hi 5	s His	Gln			10					15		
<211> 587 <212> PRT <213> Aral <400> 46 Met Lys Arg 1	Asp Hi 5 Ser Se 20	s His	Gln Ser	Ile	Ser 25	10 Lys	Asp	Lys	Met	Met 30	15 Met	Val	
<pre>&lt;211&gt; 587 &lt;212&gt; PRT &lt;213&gt; Aral &lt;400&gt; 46 Met Lys Arg 1 Thr Ser Ser Lys Lys Glo</pre>	Asp Hi 5 Ser Se 20	s His r Ser p Gly	Gln Ser Gly	Ile Gly 40	Ser 25 Asn	10 Lys Met	Asp	Lys	Met Glu 45	Met 30 Leu	15 Met Leu	Val Ala	
<pre>&lt;211&gt; 587 &lt;212&gt; PRT &lt;213&gt; Aral &lt;400&gt; 46 Met Lys Arg 1 Thr Ser Ser Lys Lys Glu 35</pre>	Asp Hi 5 Ser Se 20 Glu As	s His r Ser p Gly s Val	Gln Ser Gly Arg 55	Ile Gly 40	Ser 25 Asn Ser	Lys Met	Asp Asp Met	Lys Asp Ala 60	Met Glu 45 Glu	Met 30 Leu Val	15 Met Leu Ala	Val Ala Leu	
<pre>&lt;211&gt; 587 &lt;212&gt; PRT &lt;213&gt; Aral &lt;400&gt; 46  Met Lys Arg 1  Thr Ser Ser  Lys Lys Glu 35  Val Leu Gly 50</pre>	Asp Hi 5 Ser Ser 20 As Tyr Ly	s His r Ser p Gly s Val	Gln Ser Gly Arg 55 Thr	Ile Gly 40 Ser Met	Ser 25 Asn Ser Met	Lys Met Glu Ser	Asp Asp Met Asn 75	Lys Asp Ala 60 Val	Met Glu 45 Glu Gln	Met 30 Leu Val Glu	Met Leu Ala Asp	Val Ala Leu Gly 80	

Page 60

Pro Ala Ser Ser Asn Gly Leu Asp Pro Val Leu Pro Ser Pro Glu Ile 115 120 125

mbi19 Sequence Listing.ST25

Cys Gly Phe Pro Ala Ser Asp Tyr Asp Leu Lys Val Ile Pro Gly Asn
130 135 140

Ala Ile Tyr Gln Phe Pro Ala Ile Asp Ser Ser Ser Ser Ser Asn Asn 145 150 150 160

Gln Asn Lys Arg Leu Lys Ser Cys Ser Ser Pro Asp Ser Met Val Thr 165 170 175

Ser Thr Ser Thr Gly Thr Gln Ile Gly Gly Val Ile Gly Thr Thr Val

Thr Thr Thr Thr Thr Thr Thr Ala Ala Ala Glu Ser Thr Arg Ser 195 200 205

Val Ile Leu Val Asp Ser Gln Glu Asn Gly Val Arg Leu Val His Ala 210 215 220

Leu Met Ala Cys Ala Glu Ala Ile Gln Gln Asn Asn Leu Thr Leu Ala 225 230 240

Glu Ala Leu Val Lys Gln Ile Gly Cys Leu Ala Val Ser Gln Ala Gly 245 250 255

Ala Met Arg Lys Val Ala Thr Tyr Phe Ala Glu Ala Leu Ala Arg Arg 260 265 270

Ile Tyr Arg Leu Ser Pro Pro Gln Asn Gln Ile Asp His Cys Leu Ser 275 280 285

Asp Thr Leu Gln Met His Phe Tyr Glu Thr Cys Pro Tyr Leu Lys Phe 290 295 300

Ala His Phe Thr Ala Asn Gln Ala Ile Leu Glu Ala Phe Glu Gly Lys 305 310 315 320

Lys Arg Val His Val Ile Asp Phe Ser Met Asn Gln Gly Leu Gln Trp 325 330 335

Pro Ala Leu Met Gln Ala Leu Ala Leu Arg Glu Gly Gly Pro Pro Thr 340 345 350

Phe Arg Leu Thr Gly Ile Gly Pro Pro Ala Pro Asp Asn Ser Asp His 355 360 365

Leu His Glu Val Gly Cys Lys Leu Ala Gln Leu Ala Glu Ala Ile His  $370 \hspace{1cm} 375 \hspace{1cm} 380$ 

Val Glu Phe Glu Tyr Arg Gly Phe Val Ala Asn Ser Leu Ala Asp Leu 385 390 395 400

Asp Ala Ser Met Leu Glu Leu Arg Pro Ser Asp Thr Glu Ala Val Ala 405 415

Val Asn Ser Val Phe Glu Leu His Lys Leu Leu Gly Arg Pro Gly Gly
420 425 430

PCT/US00/31414 WO 01/35725

mbil9 Sequence Listing.ST25

Ile Glu Lys Val Leu Gly Val Val Lys Gln Ile Lys Pro Val Ile Phe Thr Val Val Glu Gln Glu Ser Asn His Asn Gly Pro Val Phe Leu Asp · 455 Arg Phe Thr Glu Ser Leu His Tyr Tyr Ser Thr Leu Phe Asp Ser Leu

Glu Gly Val Pro Asn Ser Gln Asp Lys Val Met Ser Glu Val Tyr Leu

Gly Lys Gln Ile Cys Asn Leu Val Ala Cys Glu Gly Pro Asp Arg Val

Glu Arg His Glu Thr Leu Ser Gln Trp Gly Asn Arg Phe Gly Ser Ser

Gly Leu Ala Pro Ala His Leu Gly Ser Asn Ala Phe Lys Gln Ala Ser

Met Leu Leu Ser Val Phe Asn Ser Gly Gln Gly Tyr Arg Val Glu Glu

Ser Asn Gly Cys Leu Met Leu Gly Trp His Thr Arg Pro Leu Ile Thr

Thr Ser Ala Trp Lys Leu Ser Thr Ala Ala His

<210> 47 <211> 990 <212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS <222> (7)..(813)

<223> G529

<400> 47 gcaaag atg tct tct gat tcg tcc cgg gaa gag aat gtg tac ttg gcc Met Ser Ser Asp Ser Ser Arg Glu Glu Asn Val Tyr Leu Ala 48 aag tta gcc gag caa gct gag cgt tac gag gaa atg gtt gag ttc atg Lys Leu Ala Glu Gln Ala Glu Arg Tyr Glu Glu Met Val Glu Phe Met 96 gag aaa gtt gca aag acc gtg gag acc gag gaa ctt act gtt gaa gag Glu Lys Val Ala Lys Thr Val Glu Thr Glu Glu Leu Thr Val Glu Glu agg aat ctc ttg tct gtt gct tac aag aac gtg att ggt gct agg aga Arg Asn Leu Leu Ser Val Ala Tyr Lys Asn Val Ile Gly Ala Arg Arg 50 55 60192 gct tct tgg agg att atc tct tcc att gag cag aag gaa gat agc agg Ala Ser Trp Arg Ile Ile Ser Ser Ile Glu Gln Lys Glu Asp Ser Arg 240 288 ggc aac agt gat cat gtt tcg att atc aag gat tac aga ggc aag att Gly Asn Ser Asp His Val Ser Ile Ile Lys Asp Tyr Arg Gly Lys Ile

Page 62

•	100	11331	23													101/03
	80					85	mbi	19 S	eque	nce :	List 90	ing.	ST25			
gaa Glu 95	act Thr	gag Glu	ctc Leu	agc Ser	aag Lys 100	att Ile	tgt Cys	gat Asp	ggc Gly	att Ile 105	ttg Leu	aac Asn	ctt Leu	ctt Leu	gag Glu 110	336
gct Ala	cat His	ctc Leu	att Ile	cct Pro 115	gct Ala	gct Ala	tct Ser	ttg Leu	gct Ala 120	gag Glu	tcc Ser	aaa Lys	gtt Val	ttt Phe 125	tac Tyr	384
ctg Leu	aag Lys	atg Met	aag Lys 130	gga Gly	gat Asp	tat Tyr	cat His	cgg Arg 135	tac Tyr	ctt Leu	gct Ala	gaa Glu	ttc Phe 140	aag Lys	act Thr	432
ggt Gly	gct Ala	gag Glu 145	agg Arg	aaa Lys	gaa Glu	gct Ala	gct Ala 150	gag Glu	agc Ser	act Thr	ctt Leu	gtt Val 155	gcc Ala	tac Tyr	aag Lys	480
tct Ser	gct Ala 160	cag Gln	gat Asp	att Ile	gct Ala	ctt Leu 165	gct Ala	gat Asp	ctg Leu	gct Ala	ccc Pro 170	act Thr	cac His	cca Pro	atc Ile	528
aga Arg 175	ctg Leu	999 Gly	ctt Leu	gct Ala	ctt Leu 180	aac Asn	ttc Phe	tct Ser	gtt Val	ttc Phe 185	tac Tyr	tat Tyr	gag Glu	att Ile	ctc Leu 190	576
aac Asn	tca Ser	tct Ser	gat Asp	cgt Arg 195	gcg Ala	tgt Cys	agt Ser	ctc Leu	gca Ala 200	aag Lys	cag Gln	gct Ala	ttt Phe	gat Asp 205	gag Glu	624
gca Ala	atc Ile	tcg Ser	gag Glu 210	cta Leu	gac Asp	aca Thr	ttg Leu	gga Gly 215	gag Glu	gaa Glu	tca Ser	tac Tyr	aag Lys 220	gac Asp	agt Ser	. 672
aca Thr	ttg Leu	atc Ile 225	atg Met	cag Gln	ctt Leu	ctc Leu	cgt Arg 230	gac Asp	aat Asn	ctc Leu	acc Thr	ctc Leu 235	tgg Trp	act Thr	tct Ser	720
gac Asp	ctc Leu 240	aat Asn	gac Asp	gaa Glu	gct Ala	ggt Gly 245	gat Asp	gat Asp	atc Ile	aag Lys	gaa Glu 250	gcc Ala	ccg Pro	aaa Lys	gag Glu	768
gtg Val 255	cag Gln	aaa Lys	gtt Val	gat Asp	gaa Glu 260	caa Gln	gcc Ala	caa Gln	cca Pro	cca Pro 265	cct Pro	tcg Ser	cag Gln	tga		813
taaa	atca	aga t	cca	ggaa	at ga	attt	gcaga	a caa	aaaa	gata	tate	gcti	gg t	ttct	gtgttt	873
ttaa	acaç	gaa a	aaaa	accti	tg ta	agtti	cct	t aaa	acat	gggc	tgta	agtti	cc t	taa	acatgg	933
atti	gtag	gta g	gtaat	tgta	ag ct	gcat	gati	t tg	gttai	cga	tgg	taa	aaa a	aaaa	aaa	990
<210 <211 <212 <213	L> : 2> 1	18 268 PRT Arab:	idops	sis (	thal:	iana										
<400	)> 4	48														
Mot	Cor	202	non.	Car	Car	Δτα	Glu	Glu	Agn	Va l	Tvr	Leu	Ala	Lvs	Leu	

Met Ser Ser Asp Ser Ser Arg Glu Glu Asn Val Tyr Leu Ala Lys Leu 1 10 15

Ala Glu Gln Ala Glu Arg Tyr Glu Glu Met Val Glu Phe Met Glu Lys 20 25 30

Val Ala Lys Thr Val Glu Thr Glu Glu Leu Thr Val Glu Glu Arg Asn 35 45

Leu Leu Ser Val Ala Tyr Lys Asn Val Ile Gly Ala Arg Arg Ala Ser 50 55 60

mbi19 Sequence Listing.ST25

Trp Arg Ile Ile Ser Ser Ile Glu Gln Lys Glu Asp Ser Arg Gly Asn

Ser Asp His Val Ser Ile Ile Lys Asp Tyr Arg Gly Lys Ile Glu Thr

Glu Leu Ser Lys Ile Cys Asp Gly Ile Leu Asn Leu Leu Glu Ala His 100 105 110

Leu Ile Pro Ala Ala Ser Leu Ala Glu Ser Lys Val Phe Tyr Leu Lys

Met Lys Gly Asp Tyr His Arg Tyr Leu Ala Glu Phe Lys Thr Gly Ala 130 135 140

Glu Arg Lys Glu Ala Ala Glu Ser Thr Leu Val Ala Tyr Lys Ser Ala

Gln Asp Ile Ala Leu Ala Asp Leu Ala Pro Thr His Pro Ile Arg Leu

Gly Leu Ala Leu Asn Phe Ser Val Phe Tyr Tyr Glu Ile Leu Asn Ser

Ser Asp Arg Ala Cys Ser Leu Ala Lys Gln Ala Phe Asp Glu Ala Ile

Ser Glu Leu Asp Thr Leu Gly Glu Glu Ser Tyr Lys Asp Ser Thr Leu 210 215 220

Ile Met Gln Leu Leu Arg Asp Asn Leu Thr Leu Trp Thr Ser Asp Leu

Asn Asp Glu Ala Gly Asp Asp Ile Lys Glu Ala Pro Lys Glu Val Gln 245 250 255

Lys Val Asp Glu Gln Ala Gln Pro Pro Pro Ser Gln

<210>

<211> 1069 <212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (70)..(867) <223> G531

<400> 49 ccacgcgtcc gcgacaaatc cctaattcca aaatcctctc ggatctaaaa gagtgtgtga

111

60

taacataag atg tcg tct tct cgg gaa gag aat gtg tac tta gcc aag tta Met Ser Ser Ser Arg Glu Glu Asn Val Tyr Leu Ala Lys Leu 1 5  $^{\circ}$ 

gct gag caa gct gaa cgt tat gag gaa atg gtt gag ttc atg gag aaa Ala Glu Gln Ala Glu Arg Tyr Glu Glu Met Val Glu Phe Met Glu Lys 25

159

PCT/US00/31414 WO 01/35725

												,				
							gat	gag	ctt	act		gaa	gag Glu			20
ctc Leu	ttg Leu	tct Ser	gtt Val 50	gct Ala	tac Tyr	aag Lys	aac Asn	gtc Val 55	att Ile	ggt Gly	gct Ala	agg Arg	aga Arg 60	gct Ala	tcc Ser	25
tgg Trp	agg Arg	atc Ile 65	ata Ile	tct Ser	tcc Ser	att Ile	gaa Glu 70	cag Gln	aag Lys	gaa Glu	gaa Glu	agc Ser 75	aga Arg	gga Gly	aac Asn	30:
gat Asp	gat Asp 80	cat His	gtt Val	tcc Ser	att Ile	atc Ile 85	aag Lys	gac Asp	tac Tyr	aga Arg	gga Gly 90	aag Lys	atc Ile	gaa Glu	act Thr	35:
gaa Glu 95	ctc Leu	agc Ser	aaa Lys	atc Ile	tgt Cys 100	gat Asp	gga Gly	ata Ile	ctc Leu	aat Asn 105	ctt Leu	ctg Leu	gat Asp	tct Ser	cac His 110	399
ctt Leu	gtt Val	ccc Pro	act Thr	gca Ala 115	tct Ser	ttg Leu	gcc Ala	gag Glu	tcc Ser 120	aaa Lys	gtc Val	ttt Phe	tac Tyr	ctc Leu 125	aaa Lys	44
atg Met	aaa Lys	gga Gly	gat Asp 130	tac Tyr	cac His	agg Arg	tac Tyr	ctt Leu 135	gct Ala	gag Glu	ttt Phe	aag Lys	act Thr 140	gga Gly	gct Ala	495
gag Glu	agg Arg	aaa Lys 145	gaa Glu	gct Ala	gct Ala	gag Glu	agc Ser 150	act Thr	ctg Leu	gtt Val	gct Ala	tac Tyr 155	aag Lys	tca Ser	gct Ala	543
cag Gln	gat Asp 160	att Ile	gca Ala	ctt Leu	gct Ala	gat Asp 165	tta Leu	gct Ala	cct Pro	act Thr	cat His 170	ccg Pro	att Ile	aga Arg	ctg Leu	591
gga Gly 175	ctt Leu	gct Ala	ctt Leu	aac Asn	ttc Phe 180	tct Ser	gtc Val	ttc Phe	tac Tyr	tac Tyr 185	gag Glu	att Ile	ctc Leu	aac Asn	tca Ser 190	639
cct Pro	gat Asp	cgt Arg	gcc Ala	tgc Cys 195	agt Ser	ctc Leu	gca Ala	aaa Lys	cag Gln 200	gct Ala	ttt Phe	gat Asp	gag Glu	gcc Ala 205	att Ile	687
tct Ser	gag Glu	ctg Leu	gat Asp 210	aca Thr	tta Leu	gga Gly	gaa Glu	gaa Glu 215	tca Ser	tac Tyr	aaa Lys	gac Asp	agt Ser 220	acg Thr	ttg Leu	735
ata Ile	atg Met	caa Gln 225	ctt Leu	ctc Leu	cgt Arg	gac Asp	aat Asn 230	ctg Leu	acc Thr	ctt Leu	tgg Trp	aac Asn 235	tct Ser	gac Asp	atc Ile	783
aat Asn	gat Asp 240	gag Glu	gcg Ala	ggc Gly	ggt Gly	gat Asp 245	gag Glu	atc Ile	aag Lys	gag Glu	gcg Ala 250	tca Ser	aaa Lys	cat His	gag Glu	831
ccg Pro 255	gaa Glu	gag Glu	999 Gly	aaa Lys	cca Pro 260	gct Ala	gag Glu	aca Thr	999 Gly	cag Gln 265	tga	ccag	gagag	gag		877
aggg	gagag	gac a	ittt	taag	gt at	gtat	ggta	a tgg	gattt	tcc	aaaa	acat	at	gatat	gatta	937
9999	attt	gt a	ıgaaç	gcaga	ag ag	gaaag	gatet	: tta	itttg	att	ttct	ccaa	aaa a	atcto	etgttc	997
															aaaaa	1057
	aaaa															1069

<400> 50

<210> 50 <211> 265 <212> PRT <213> Arabidopsis thaliana

PCT/US00/31414 WO 01/35725

### mbi19 Sequence Listing.ST25

Met Ser Ser Ser Arg Glu Glu Asn Val Tyr Leu Ala Lys Leu Ala Glu 1 5 10 15

Gln Ala Glu Arg Tyr Glu Glu Met Val Glu Phe Met Glu Lys Val Ala 20 25 30

Lys Thr Val Asp Thr Asp Glu Leu Thr Val Glu Glu Arg Asn Leu Leu

Ser Val Ala Tyr Lys Asn Val Ile Gly Ala Arg Arg Ala Ser Trp Arg 50 55 60

Ile Ile Ser Ser Ile Glu Gln Lys Glu Glu Ser Arg Gly Asn Asp Asp 65 70 75 80

His Val Ser Ile Ile Lys Asp Tyr Arg Gly Lys Ile Glu Thr Glu Leu

Ser Lys Ile Cys Asp Gly Ile Leu Asn Leu Leu Asp Ser His Leu Val 100 105 110

Pro Thr Ala Ser Leu Ala Glu Ser Lys Val Phe Tyr Leu Lys Met Lys

Gly Asp Tyr His Arg Tyr Leu Ala Glu Phe Lys Thr Gly Ala Glu Arg

Lys Glu Ala Ala Glu Ser Thr Leu Val Ala Tyr Lys Ser Ala Gln Asp 145 150 160

Ile Ala Leu Ala Asp Leu Ala Pro Thr His Pro Ile Arg Leu Gly Leu 165 170 175

Ala Leu Asn Phe Ser Val Phe Tyr Tyr Glu Ile Leu Asn Ser Pro Asp

Arg Ala Cys Ser Leu Ala Lys Gln Ala Phe Asp Glu Ala Ile Ser Glu 195 200 205

Leu Asp Thr Leu Gly Glu Glu Ser Tyr Lys Asp Ser Thr Leu Ile Met 210 215 220

Gln Leu Leu Arg Asp Asn Leu Thr Leu Trp Asn Ser Asp Ile Asn Asp

Glu Ala Gly Gly Asp Glu Ile Lys Glu Ala Ser Lys His Glu Pro Glu 245 250 255

Glu Gly Lys Pro Ala Glu Thr Gly Gln

<210> 51 <211> 2240

<212> DNA <213> Arabidopsis thaliana

<220>

### mbil9 Sequence Listing.ST25

<221> CDS <222> (238)..(2064) <223> G214

<400> 51 60 tgagatttct ccatttccgt agcttctggt ctcttttctt tgtttcattg atcaaaagca 120 aatcacttct tottottctt ottotogatt tottactgtt ttottatcca acgaaatctg gaattaaaaa tggaatcttt atcgaatcca agctgatttt gtttctttca ttgaatcatc 180 tctctaaagt ggaattttgt aaagagaaga tctgaagttg tgtagaggag cttagtg 237 atg gag aca aat tcg tct gga gaa gat ctg gtt att aag act cgg aag Met Glu Thr Asn Ser Ser Gly Glu Asp Leu Val Ile Lys Thr Arg Lys 285 cca tat acg ata aca aag caa cgt gaa agg tgg act gag gaa gaa cat Pro Tyr Thr Ile Thr Lys Gln Arg Glu Arg Trp Thr Glu Glu Glu His 333 aat aga ttc att gaa gct ttg agg ctt tat ggt aga gca tgg cag aag Asn Arg Phe Ile Glu Ala Leu Arg Leu Tyr Gly Arg Ala Trp Gln Lys 381 att gaa gaa cat gta gca aca aaa act gct gtc cag ata aga agt cac Ile Glu Glu His Val Ala Thr Lys Thr Ala Val Gln Ile Arg Ser His gct cag aaa ttt ttc tcc aag gta gag aaa gag gct gaa gct aaa ggt Ala Gln Lys Phe Phe Ser Lys Val Glu Lys Glu Ala Glu Ala Lys Gly 477 gta gct atg ggt caa gcg cta gac ata gct att cct cct cca cgg cct Val Ala Met Gly Gln Ala Leu Asp Ile Ala Ile Pro Pro Pro Arg Pro 525 573 aag cgt aaa cca aac aat cct tat cct cga aag acg gga agt gga acg Lys Arg Lys Pro Asn Asn Pro Tyr Pro Arg Lys Thr Gly Ser Gly Thr atc ctt atg tca aaa acg ggt gtg aat gat gga aaa gag tcc ctt gga Ile Leu Met Ser Lys Thr Gly Val Asn Asp Gly Lys Glu Ser Leu Gly 115 120 125 621 tca gaa aaa gtg tcg cat cct gag atg gcc aat gaa gat cga caa caa Ser Glu Lys Val Ser His Pro Glu Met Ala Asn Glu Asp Arg Gln Gln 669 tca aag cct gaa gag aaa act ctg cag gaa gac aac tgt tca gat tgt Ser Lys Pro Glu Glu Lys Thr Leu Gln Glu Asp Asn Cys Ser Asp Cys 717 ttc act cat cag tat ctc tct gct gca tcc tcc atg aat aaa agt tgt Phe Thr His Gln Tyr Leu Ser Ala Ala Ser Ser Met Asn Lys Ser Cys 765 ata gag aca tca aac gca agc act ttc cgc gag ttc ttg cct tca cgg Ile Glu Thr Ser Asn Ala Ser Thr Phe Arg Glu Phe Leu Pro Ser Arg 813 gaa gag gga agt cag aat aac agg gta aga aag gag tca aac tca gat Glu Glu Gly Ser Gln Asn Asn Arg Val Arg Lys Glu Ser Asn Ser Asp 861 ttg aat gca aaa tct ctg gaa aac ggt aat gag caa gga cct cag act 909 Leu Asn Ala Lys Ser Leu Glu Asn Gly Asn Glu Gln Gly Pro Gln Thr tat ccg atg cat atc cct gtg cta gtg cca ttg ggg agc tca ata aca Tyr Pro Met His Ile Pro Val Leu Val Pro Leu Gly Ser Ser Ile Thr 957 agt tot ota toa cat cot cot toa gag coa gat agt cat coc cac aca 1005

Page 67

Ser	Ser	Leu	Ser	His 245	Pro	Pro		19 S Glu							Thr	
gtt Val	gca Ala	gga Gly	gat Asp 260	tat Tyr	cag Gln	tcg Ser	ttt Phe	cct Pro 265	aat Asn	cat His	ata Ile	atg Met	tca Ser 270	acc Thr	ctt Leu	1053
tta Leu	caa Gln	aca Thr 275	ccg Pro	gct Ala	ctt Leu	tat Tyr	act Thr 280	gcc Ala	gca Ala	act Thr	ttc Phe	gcc Ala 285	tca Ser	tca Ser	ttt Phe	1101
tgg Trp	cct Pro 290	ccc Pro	gat Asp	tct Ser	agt Ser	ggt Gly 295	ggc Gly	tca Ser	cct Pro	gtt Val	cca Pro 300	999 Gly	aac Asn	tca Ser	cct Pro	1149
ccg Pro 305	aat Asn	ctg Leu	gct Ala	gcc Ala	atg Met 310	gcc Ala	gca Ala	gcc Ala	act Thr	gtt Val 315	gca Ala	gct Ala	gct Ala	agt Ser	gct Ala 320	1197
tgg Trp	tgg Trp	gct Ala	gcc Ala	aat Asn 325	gga Gly	tta Leu	tta Leu	cct Pro	tta Leu 330	tgt Cys	gct Ala	cct Pro	ctt Leu	agt Ser 335	tca Ser	1245
ggt Gly	ggt Gly	ttc Phe	act Thr 340	agt Ser	cat His	cct Pro	cca Pro	tct Ser 345	act Thr	ttt Phe	gga Gly	cca Pro	tca Ser 350	tgt Cys	gat Asp	1293
						agc Ser										1341
						gag Glu 375										1389
tca Ser 385	gag Glu	gat Asp	gtt Val	gaa Glu	aat Asn 390	aag Lys	agt Ser	aaa Lys	cca Pro	gtt Val 395	tgt Сув	cat His	gag Glu	cag Gln	cct Pro 400	1437
						gat Asp										1485
						tcc Ser										1533
agt Ser	gat Asp	gat Asp 435	gtt Val	gag Glu	gcg Ala	gat Asp	gca Ala 440	tca Ser	gaa Glu	agg Arg	caa Gln	gag Glu 445	gat Asp	ggc Gly	acc Thr	1581
aat Asn	ggt Gly 450	gag Glu	gtg Val	aaa Lys	gaa Glu	acg Thr 455	aat Asn	gaa Glu	gac Asp	act Thr	aat Asn 460	aaa Lys	cct Pro	caa Gln	act Thr	1629
tca Ser 465	gag Glu	tcc Ser	aat Asn	gca Ala	cgc Arg 470	cgc Arg	agt Ser	aga Arg	atc Ile	agc Ser 475	tcc Ser	aat Asn	ata Ile	acc Thr	gat Asp 480	. 1677
						gac Asp										1725
ttc Phe	tcc Ser	aga Arg	gag Glu 500	gta Val	ttg Leu	ccg Pro	caa Gln	agt Ser 505	ttt Phe	aca Thr	tat Tyr	cga Arg	gaa Glu 510	gaa Glu	cac His	1773
						caa Gln										1821
gat Asp	ctt Leu 530	aac Asn	ttc Phe	aca Thr	gct Ala	cag Gln 535	tta Leu	aca Thr	cca Pro	gtt Val	gat Asp 540	gat Asp	caa Gln	gag Glu	gag Glu	1869

aaq aqa aac aca qqa ttt ctt qq	oil9 Sequence Listing.ST25 a atc gga tta gat gct tca aag cta y Ile Gly Leu Asp Ala Ser Lys Leu 555 560	1917
atg agt aga gga aga aca ggt tt Met Ser Arg Gly Arg Thr Gly Ph 565	t aaa cca tac aaa aga tgt tcc atg te Lys Pro Tyr Lys Arg Cys Ser Met 570 575	1965
gaa gcc aaa gaa agt aga atc ct Glu Ala Lys Glu Ser Arg Ile Le 580	c aac aac aat cct atc att cat gtg eu Asn Asn Asn Pro Ile Ile His Val 585 590	2013
gaa cag aaa gat ccc aaa cgg at Glu Gln Lys Asp Pro Lys Arg Me 595 60	g cgg ttg gaa act caa gct tcc aca t Arg Leu Glu Thr Gln Ala Ser Thr 605	2061
tga gactctattt tcatctgatc tgtt	gtttgt actctgtttt taagttttca	2114
agaccactgc tacattttct ttttcttt	tg aggeetttgt atttgtttee ttgteeatag	2174
	itt caacaaatca taaactgttt aatctttttt	2234
tttcca		2240
<210> 52 <211> 608 <212> PRT <213> Arabidopsis thaliana		
<400> 52		
Met Glu Thr Asn Ser Ser Gly Gl	u Asp Leu Val Ile Lys Thr Arg Lys 10 15	
Pro Tyr Thr Ile Thr Lys Gln Ar 20	rg Glu Arg Trp Thr Glu Glu Glu His 25 30	
Asn Arg Phe Ile Glu Ala Leu Ar 35 40	rg Leu Tyr Gly Arg Ala Trp Gln Lys 45	
Ile Glu Glu His Val Ala Thr Ly 50 55	rs Thr Ala Val Gln Ile Arg Ser His 60	
Ala Gln Lys Phe Phe Ser Lys Va	ıl Glu Lys Glu Ala Glu Ala Lys Gly 75 80	
Val Ala Met Gly Gln Ala Leu As 85	sp Ile Ala Ile Pro Pro Pro Arg Pro 90 95	
Lys Arg Lys Pro Asn Asn Pro Ty 100	or Pro Arg Lys Thr Gly Ser Gly Thr 105 110	
Ile Leu Met Ser Lys Thr Gly Va 115 12	al Asn Asp Gly Lys Glu Ser Leu Gly 20 125	
Ser Glu Lys Val Ser His Pro Gl 130 135	u Met Ala Asn Glu Asp Arg Gln Gln 140	
Ser Lys Pro Glu Glu Lys Thr Le 145 150	eu Gln Glu Asp Asn Cys Ser Asp Cys 155 . 160	
Phe Thr His Gln Tyr Leu Ser Al 165	a Ala Ser Ser Met Asn Lys Ser Cys 170 175	

mbil9 Sequence Listing.ST25

Ile Glu Thr Ser Asn Ala Ser Thr Phe Arg Glu Phe Leu Pro Ser Arg

Glu Glu Gly Ser Gln Asn Asn Arg Val Arg Lys Glu Ser Asn Ser Asp 195 200 205

Leu Asn Ala Lys Ser Leu Glu Asn Gly Asn Glu Gln Gly Pro Gln Thr 210 215 220

Tyr Pro Met His Ile Pro Val Leu Val Pro Leu Gly Ser Ser Ile Thr 225 230 235 240

Ser Ser Leu Ser His Pro Pro Ser Glu Pro Asp Ser His Pro His Thr 245 250 255

Val Ala Gly Asp Tyr Gln Ser Phe Pro Asn His Ile Met Ser Thr Leu 260 265 270

Leu Gln Thr Pro Ala Leu Tyr Thr Ala Ala Thr Phe Ala Ser Ser Phe  $\frac{275}{280}$ 

Trp Pro Pro Asp Ser Ser Gly Gly Ser Pro Val Pro Gly Asn Ser Pro 290 295 300

Pro Asn Leu Ala Ala Met Ala Ala Ala Thr Val Ala Ala Ala Ser Ala 305 310 315 320

Trp Trp Ala Ala Asn Gly Leu Leu Pro Leu Cys Ala Pro Leu Ser Ser 325 330 335

Gly Gly Phe Thr Ser His Pro Pro Ser Thr Phe Gly Pro Ser Cys Asp 340 345 350

Val Glu Tyr Thr Lys Ala Ser Thr Leu Gln His Gly Ser Val Gln Ser

Arg Glu Glu His Ser Glu Ala Ser Lys Ala Arg Ser Ser Leu Asp 370 375 380

Ser Glu Asp Val Glu Asn Lys Ser Lys Pro Val Cys His Glu Gln Pro 385 390 395 400

Ser Ala Thr Pro Glu Ser Asp Ala Lys Gly Ser Asp Gly Ala Gly Asp 405 410 415

Arg Lys Gln Val Asp Arg Ser Ser Cys Gly Ser Asn Thr Pro Ser Ser 420 425 430

Ser Asp Asp Val Glu Ala Asp Ala Ser Glu Arg Gln Glu Asp Gly Thr 435 440 445

Asn Gly Glu Val Lys Glu Thr Asn Glu Asp Thr Asn Lys Pro Gln Thr 450 455 460

Ser Glu Ser Asn Ala Arg Arg Ser Arg Ile Ser Ser Asn Ile Thr Asp 465 470 475 480

## mbi19 Sequence Listing.ST25

Pro	Trp	Lys	Ser	Val 485	Ser	Авр	Glu	Gly	Arg 490	Ile	Ala	Phe	Gln	Ala 495	Leu	
Phe	Ser	Arg	Glu 500	Val	Leu	Pro	Gln	Ser 505	Phe	Thr	Tyr	Arg	Glu 510	Glu	His	
Arg	Glu	Glu 515	Glu	Gln	Gln	Gln	Gln 520	Glu	Gln	Arg	Tyr	Pro 525	Met	Ala	Leu	
Asp	Leu 530	Asn	Phe	Thr	Ala	Gln 535	Leu	Thr	Pro	Val	Asp 540	Asp	Gln	Glu	Glu	
Lys 545	Arg	Asn	Thr	Gly	Phe 550	Leu	Gly	Ile	Gly	Leu 555	Asp	Ala	Ser	Lys	Leu 560	
Met	Ser	Arg	Gly	Arg 565	Thr	Gly	Phe	Lys	Pro 570	Tyr	Lys	Arg	Сув	Ser 575	Met	
Glu	Ala	ГÀЗ	Glu 580	Ser	Arg	Ile	Leu	Asn 585	Asn	Asn	Pro	Ile	Ile 590	His	Val	
Glu	Gln	Lув 595	Asp	Pro	Lys	Arg	Met 600	Arg	Leu	Glu	Thr	Gln 605	Ala	Ser	Thr	
<210 <210 <210 <210	1> : 2> I	53 1155 DNA Arab:	idops	sis (	thali	iana										
<220 <221 <221 <221	1> ( 2>	CDS (76) 3193	(10 )	077)												
<400 atto	0> !	53 cta (	ctaat	tctci	tc aa	igati	tcac	c aat	tttc	ttg	tgat	ttt	ctc t	cagi	ttctt	60
				c ato	g gat	ge	ate	g agt	: ago	gta	a gad	gag	g ago	e te	aca Thr	111
act Thr	aca Thr	gat Asp 15	tcc Ser	att Ile	ccg Pro	gcg Ala	aga Arg 20	aag Lys	tca Ser	tcg Ser	tct Ser	ccg Pro 25	gcg Ala	agt Ser	tta Leu	159
cta Leu	tat Tyr 30	aga Arg	atg Met	gga Gly	agc Ser	gga Gly 35	aca Thr	agc Ser	gtg Val	gta Val	ctt Leu 40	gat Asp	tca Ser	gag Glu	aac Asn	207
ggt Gly 45	gtc Val	gaa Glu	gtc Val	gaa Glu	gtc Val 50	gaa Glu	gcc Ala	gaa Glu	tca Ser	aga Arg 55	aag Lys	ctt Leu	cct Pro	tct Ser	tca Ser 60	255
aga Arg	ttc Phe	aaa Lys	ggt Gly	gtt Val 65	gtt Val	cct Pro	caa Gln	cca Pro	aat Asn 70	gga Gly	aga Arg	tgg Trp	gga Gly	gct Ala 75	cag Gln	303
att	tac	gag	aaa	cat	caa	cgc	gtg Val	tgg Trp	ctt Leu	ggt Glv	act Thr	ttc Phe	aac Asn	gag Glu	gaa Glu	351
	Tyr	Giu	80	HIS	GIII			85		2			90			
gac	Tyr gaa Glu	qca	80 act	cat	act	tac	qac	85 qtc	gcg Ala	qct	cac His	cgt	90 ttc	cgt	ggc	399

Page 71

PCT/US00/31414 WO 01/35725

## mbi19 Sequence Listing.ST25

									-			-				
cgc Arg	gat Asp 110	gcc Ala	gtt Val	act Thr	aat Asn	ttc Phe 115	aaa Lys	gac Asp	acg Thr	acg Thr	ttc Phe 120	gaa Glu	gaa Glu	gag Glu	gtt Val	447
gag Glu 125	ttc Phe	tta Leu	aac Asn	gcg Ala	cat His 130	tcg Ser	aaa Lys	tca Ser	gag Glu	atc Ile 135	gta Val	gat Asp	atg Met	ttg Leu	aga Arg 140	495
aaa Lys	cac His	act Thr	tac Tyr	aaa Lys 145	gaa Glu	gag Glu	tta Leu	gac Asp	caa Gln 150	agg Arg	aaa Lys	cgt Arg	aac Asn	cgt Arg 155	gac Asp	543
ggt Gly	aac Asn	gga Gly	aaa Lys 160	gag Glu	acg Thr	acg Thr	gcg Ala	ttt Phe 165	gct Ala	ttg Leu	gct Ala	tcg Ser	atg Met 170	gtg Val	gtt Val	591
atg Met	acg Thr	999 Gly 175	ttt Phe	aaa Lys	acg Thr	gcg Ala	gag Glu 180	tta Leu	ctg Leu	ttt Phe	gag Glu	aaa Lys 185	acg Thr	gta Val	acg Thr	639
cca Pro	agt Ser 190	gac Asp	gtc Val	999 Gly	aaa Lys	cta Leu 195	aac Asn	cgt Arg	tta Leu	gtt Val	ata Ile 200	cca Pro	aaa Lys	cac His	caa Gln	687
gcg Ala 205	gag Glu	aaa Lys	cat His	ttt Phe	ccg Pro 210	tta Leu	ccg Pro	tta Leu	ggt Gly	aat Asn 215	aat Asn	aac Asn	gtc Val	tcc Ser	gtt Val 220	735
aaa Lys	ggt Gly	atg Met	ctg Leu	ttg Leu 225	aat Asn	ttc Phe	gaa Glu	gac Asp	gtt Val 230	aac Asn	ggg Gly	aaa Lys	gtg Val	tgg Trp 235	agg Arg	783
ttc Phe	cgt Arg	tac Tyr	tct Ser 240	tat Tyr	tgg Trp	aat Asn	agt Ser	agt Ser 245	caa Gln	agt Ser	tat Tyr	gtg Val	ttg Leu 250	acc Thr	aaa Lys	831
ggt Gly	tgg Trp	agt Ser 255	aga Arg	ttc Phe	gtt Val	aaa Lys	gag Glu 260	aag Lys	aga Arg	ctt Leu	tgt Cys	gct Ala 265	ggt Gly	gat Asp	ttg Leu	879
atc Ile	agt Ser 270	ttt Phe	aaa Lys	aga Arg	tcc Ser	aac Asn 275	gat Asp	caa Gln	gat Asp	caa Gln	aaa Lys 280	ttc Phe	ttt Phe	atc Ile	999 Gly	927
tgg Trp 285	aaa Lys	tcg Ser	aaa Lys	tcc Ser	999 Gly 290	ttg Leu	gat Asp	cta Leu	gag Glu	acg Thr 295	ggt Gly	cgg Arg	gtt Val	atg Met	aga Arg 300	975
ttg Leu	ttt Phe	gly ggg	gtt Val	gat Asp 305	att Ile	tct Ser	tta Leu	aac Asn	gcc Ala 310	gtc Val	gtt Val	gta Val	gtg Val	aag Lys 315	gaa Glu	1023
			gtg Val 320													1071
ttg Leu	taa	taad	caatt	ta a	acaad	ettgg	gg aa	aagaa	aaaa	a aaq	gcttt	ttg	atti	taat	tt	1127
ctct	tcaa	acg t	taat	ctt	gc to	gagat	ta									1155

<210> 54 <211> 333 <212> PRT <213> Arabidopsis thaliana

<400> 54

Met Asp Ala Met Ser Ser Val Asp Glu Ser Ser Thr Thr Thr Asp Ser 1  $\phantom{\Big|}$  5  $\phantom{\Big|}$  10  $\phantom{\Big|}$  15

mbi19 Sequence Listing.ST25

Ile Pro Ala Arg Lys Ser Ser Ser Pro Ala Ser Leu Leu Tyr Arg Met
20 25 30 Gly Ser Gly Thr Ser Val Val Leu Asp Ser Glu Asn Gly Val Glu Val 40 45Glu Val Glu Ala Glu Ser Arg Lys Leu Pro Ser Ser Arg Phe Lys Gly 50 60 Val Val Pro Gln Pro Asn Gly Arg Trp Gly Ala Gln Ile Tyr Glu Lys 65 70 75 His Gln Arg Val Trp Leu Gly Thr Phe Asn Glu Glu Asp Glu Ala Ala 85 90 95 Arg Ala Tyr Asp Val Ala Ala His Arg Phe Arg Gly Arg Asp Ala Val Thr Asn Phe Lys Asp Thr Thr Phe Glu Glu Glu Val Glu Phe Leu Asn Ala His Ser Lys Ser Glu Ile Val Asp Met Leu Arg Lys His Thr Tyr 130 135 140 Lys Glu Glu Leu Asp Gln Arg Lys Arg Asn Arg Asp Gly Asn Gly Lys 145 150 160 Glu Thr Thr Ala Phe Ala Leu Ala Ser Met Val Val Met Thr Gly Phe 165 170 175 Lys Thr Ala Glu Leu Leu Phe Glu Lys Thr Val Thr Pro Ser Asp Val Gly Lys Leu Asn Arg Leu Val Ile Pro Lys His Gln Ala Glu Lys His Phe Pro Leu Pro Leu Gly Asn Asn Val Ser Val Lys Gly Met Leu 210 215 220 Leu Asn Phe Glu Asp Val Asn Gly Lys Val Trp Arg Phe Arg Tyr Ser 225 230 230 235 Tyr Trp Asn Ser Ser Gln Ser Tyr Val Leu Thr Lys Gly Trp Ser Arg Phe Val Lys Glu Lys Arg Leu Cys Ala Gly Asp Leu Ile Ser Phe Lys 260 265 270 Arg Ser Asn Asp Gln Asp Gln Lys Phe Phe Ile Gly Trp Lys Ser Lys 275 280 285

Ser Gly Leu Asp Leu Glu Thr Gly Arg Val Met Arg Leu Phe Gly Val 290 295 300

Asp Ile Ser Leu Asn Ala Val Val Val Val Lys Glu Thr Thr Glu Val 305 310 315 320

## mbi19 Sequence Listing.ST25

<210> 55 <211> 1246 <212> DNA <213> Arabidopsis thaliana
<220> <221> CDS <222> (81)(1139) <223> G9
<400> 55 gtgtttcttc tttctgctaa aaggttataa tttttgtttc ttggtttggt
aagaaactga aacaaagaaa atg gat tct agt tgc ata gac gag ata agt tcc  Met Asp Ser Ser Cys Ile Asp Glu Ile Ser Ser  1 5 10
tcc act tca gaa tct ttc tcc gcc acc acc gcc aag aag ctc tct cct Ser Thr Ser Glu Ser Phe Ser Ala Thr Thr Ala Lys Lys Leu Ser Pro 15 20 25
cct ccc gcg gcg gcg tta cgc ctc tac cgg atg gga agc ggc ggg agc Pro Pro Ala Ala Ala Leu Arg Leu Tyr Arg Met Gly Ser Gly Gly Ser 30 35 40
agc gtc gtg ttg gat ccc gag aac ggc cta gag acg gag tca cga aag Ser Val Val Leu Asp Pro Glu Asn Gly Leu Glu Thr Glu Ser Arg Lys 45 50 55
Cta cca tct tca aaa tac aaa ggt gtt gtt cct cag cct aac gga aga Leu Pro Ser Ser Lys Tyr Lys Gly Val Val Pro Gln Pro Asn Gly Arg 60 65 70 75
tgg gga gct cag atc tac gag aag cac caa cga gta tgg ctc ggg act Trp Gly Ala Gln Ile Tyr Glu Lys His Gln Arg Val Trp Leu Gly Thr 80 85 90
ttc aac gag caa gaa gat gct cgt tcc tac gac atc gca gct tgt Phe Asn Glu Gln Glu Ala Ala Arg Ser Tyr Asp Ile Ala Ala Cys 95 100 105
aga ttc cgt ggc cgc gac gcc gtc gtc aac ttc aag aac gtt ctg gaa Arg Phe Arg Gly Arg Asp Ala Val Val Asn Phe Lys Asn Val Leu Glu 110 115 120
gac ggc gat tta gct ttt ctt gaa gct cac tca aag gcc gag atc gtc Asp Gly Asp Leu Ala Phe Leu Glu Ala His Ser Lys Ala Glu Ile Val 125 130 135
gac atg ttg aga aaa cac act tac gcc gac gag ctt gaa cag aac aat Asp Met Leu Arg Lys His Thr Tyr Ala Asp Glu Leu Glu Gln Asn Asn 140 155 150 155
aaa cgg cag ttg ttt ctc tcc gtc gac gct aac gga aaa cgt aac gga Lys Arg Gln Leu Phe Leu Ser Val Asp Ala Asn Gly Lys Arg Asn Gly 160 165 170
tcg agt act act caa aac gac aaa gtt tta aag acg tgt gaa gtt ctt Ser Ser Thr Thr Gln Asn Asp Lys Val Leu Lys Thr Cys Glu Val Leu 175 180 185
ttc gag aag gct gtt aca cct agc gac gtt ggg aag cta aac cgt ctc Phe Glu Lys Ala Val Thr Pro Ser Asp Val Gly Lys Leu Asn Arg Leu 190 195 200
gtg ata cct aaa caa cac gcc gag aaa cac ttt ccg tta ccg tca ccg Val Ile Pro Lys Gln His Ala Glu Lys His Phe Pro Leu Pro Ser Pro 205 210 215

Page 74

							mbi	19 S	eque	nce :	List	ing.	ST25			
tca Ser 220	ccg Pro	gca Ala	gtg Val	act Thr	aaa Lys 225	gga Gly	gtt Val	ttg Leu	atc Ile	aac Asn 230	ttc Phe	gaa Glu	gac Asp	gtt Val	aac Asn 235	785
ggt Gly	aaa Lys	gtg Val	tgg Trp	agg Arg 240	ttc Phe	cgt Arg	tac Tyr	tca Ser	tac Tyr 245	tgg Trp	aac Asn	agt Ser	agt Ser	caa Gln 250	agt Ser	833
tac Tyr	gtg Val	ttg Leu	acc Thr 255	aag Lys	gga Gly	tgg Trp	agt Ser	cga Arg 260	ttc Phe	gtc Val	aag Lys	gag Glu	aag Lys 265	aat Asn	ctt Leu	881
cga Arg	gcc Ala	ggt Gly 270	gat Asp	gtt Val	gtt Val	act Thr	ttc Phe 275	gag Glu	aga Arg	tcg Ser	acc Thr	gga Gly 280	cta Leu	gag Glu	cgg Arg	929
		tat Tyr														977
gtt Val 300	çag Gln	gtg Val	gtg Val	gtt Val	cgg Arg 305	ctt Leu	ttc Phe	gga Gly	gtt Val	gat Asp 310	atc Ile	ttt Phe	aat Asn	gtg Val	acc Thr 315	1025
acc Thr	gtg Val	aag Lys	cca Pro	aac Asn 320	gac Asp	gtc Val	gtg Val	gcc Ala	gtt Val 325	tgc Cys	ggt Gly	gga Gly	aag Lys	aga Arg 330	tct Ser	1073
cga Arg	gat Asp	gtt Val	gat Asp 335	gat Asp	atg Met	ttt Phe	gcg Ala	tta Leu 340	cgg Arg	tgt Cys	tcc Ser	aag Lys	aag Lys 345	cag Gln	gcg Ala	1121
		aat Asn 350			tga	cata	tttc	ect t	ttco	gatt	t ta	atgct	ttc	3		1169
tttt	ttaa	att t	tttt	ttt	g to	aagt	tgtg	tag	gttg	jtga	ttca	atgct	ag g	gttgt	attta	1229
ggaa	aaga	iga t	aaga	cc				-								1246
<210 <211 <211 <211	l> 3 2> I	66 152 PRT Arabi	idops	sis t	:hali	iana										
<400	)> 5	6														
Met 1	Asp	Ser	Ser	Сув 5	Ile	Asp	Glu	Ile	Ser 10	Ser	Ser	Thr	Ser	Glu 15	Ser	
Phe	Ser	Ala	Thr 20		Ala	Lys					Pro		Ala 30	Ala	Ala	
Leu	Arg	Leu 35	Tyr	Arg	Met	Gly	Ser 40	Gly	Gly	Ser	Ser	Val 45	Val	Leu	Asp	
Pro	Glu 50	Asn	Gly	Leu	Glu	Thr 55	Glu	Ser	Arg	Lys	Leu 60	Pro	Ser	Ser	Lys	
ryr 65	Lys	Gly	Val	Val	Pro 70	Gln	Pro	Asn	Gly	Arg 75	Trp	Gly	Ala	Gln	Ile 80	
Tyr	Glu	Lys	His	Gln 85	Arg	Val	Trp	Leu	Gly 90	Thr	Phe	Asn	Glu	Gln 95	Glu	
Glu	Ala	Ala	Arg 100	Ser	Tyr	Asp	Ile	Ala 105	Ala	Cys	Arg	Phe	Arg 110	Gly	Arg	

PCT/US00/31414 WO 01/35725

mbi19 Sequence Listing.ST25

Asp Ala Val Val Asn Phe Lys Asn Val Leu Glu Asp Gly Asp Leu Ala 115 120 125

Phe Leu Glu Ala His Ser Lys Ala Glu Ile Val Asp Met Leu Arg Lys 130 135 140

His Thr Tyr Ala Asp Glu Leu Glu Gln Asn Asn Lys Arg Gln Leu Phe

Leu Ser Val Asp Ala Asn Gly Lys Arg Asn Gly Ser Ser Thr Thr Gln

Asn Asp Lys Val Leu Lys Thr Cys Glu Val Leu Phe Glu Lys Ala Val

Thr Pro Ser Asp Val Gly Lys Leu Asn Arg Leu Val Ile Pro Lys Gln

His Ala Glu Lys His Phe Pro Leu Pro Ser Pro Ser Pro Ala Val Thr

Lys Gly Val Leu Ile Asn Phe Glu Asp Val Asn Gly Lys Val Trp Arg 225 230 235 240

Phe Arg Tyr Ser Tyr Trp Asn Ser Ser Gln Ser Tyr Val Leu Thr Lys 245 250 255

Gly Trp Ser Arg Phe Val Lys Glu Lys Asn Leu Arg Ala Gly Asp Val 260 265 270

Val Thr Phe Glu Arg Ser Thr Gly Leu Glu Arg Gln Leu Tyr Ile Asp

Trp Lys Val Arg Ser Gly Pro Arg Glu Asn Pro Val Gln Val Val 290 295 300

Arg Leu Phe Gly Val Asp Ile Phe Asn Val Thr Thr Val Lys Pro Asn

Asp Val Val Ala Val Cys Gly Gly Lys Arg Ser Arg Asp Val Asp Asp 325 330 335

Met Phe Ala Leu Arg Cys Ser Lys Lys Gln Ala Ile Ile Asn Ala Leu 340 350

<210> 57 <211> 1239 <212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS <222> (6)..(1091) <223> G993

<400> 57

caaat atg gaa tac agc tgt gta gac gac agt agt aca acg tca gaa tct Met Glu Tyr Ser Cys Val Asp Asp Ser Ser Thr Thr Ser Glu Ser

	1				5		mbi	19 S	eque	nce 10	List )	ing.	ST25		15	
ctc Leu	tcc Ser	atc Ile	tct Ser	act Thr 20	act Thr	cca Pro	aag Lys	ccg Pro	aca Thr 25	acg Thr	acg Thr	acg Thr	gag Glu	aag Lys 30	aaa Lys	98
ctc Leu	tct Ser	tct Ser	ccg Pro 35	ccg Pro	gcg Ala	acg Thr	tcg Ser	atg Met 40	cgt Arg	ctc Leu	tac Tyr	aga Arg	atg Met 45	gga Gly	agc Ser	. 146
ggc Gly	gga Gly	agc Ser 50	agc Ser	gtc Val	gtt Val	ttg Leu	gat Asp 55	tca Ser	gag Glu	aac Asn	ggc Gly	gtc Val 60	gag Glu	acc Thr	gag Glu	194
tca Ser	cgt Arg 65	aag Lys	ctt Leu	cct Pro	tcg Ser	tcg Ser 70	aaa Lys	tat Tyr	aaa Lys	ggc Gly	gtt Val 75	gtg Val	cct Pro	cag Gln	cct Pro	242
aac Asn 80	gga Gly	aga Arg	tgg Trp	gga Gly	gct Ala 85	cag Gln	att Ile	tac Tyr	gag Glu	aag Lys 90	cat His	cag Gln	cga Arg	gtt Val	tgg Trp 95	290
ctc Leu	ggt Gly	act Thr	ttc Phe	aac Asn 100	gag Glu	gaa Glu	gaa Glu	gaa Glu	gct Ala 105	gcg Ala	tct Ser	tct Ser	tac Tyr	gac Asp 110	atc Ile	338
gcc Ala	gtg Val	agg Arg	aga Arg 115	ttc Phe	cgc Arg	ggc Gly	cgc Arg	gac Asp 120	gcc Ala	gtc Val	act Thr	aac Asn	ttc Phe 125	aaa Lys	tct Ser	386
caa Gln	gtt Val	gat Asp 130	gga Gly	aac Asn	gac Asp	gcc Ala	gaa Glu 135	tcg Ser	gct Ala	ttt Phe	ctt Leu	gac Asp 140	gct Ala	cat His	tct Ser	434
aaa Lys	gct Ala 145	gag Glu	atc Ile	gtg Val	gat Asp	atg Met 150	ttg Leu	agg Arg	aaa Lys	cac His	act Thr 155	tac Tyr	gcc Ala	gat Asp	gag Glu	482
ttt Phe 160	gag Glu	cag Gln	agt Ser	aga Arg	cgg Arg 165	aag Lys	ttt Phe	gtt Val	aac Asn	ggc Gly 170	gac Asp	gga Gly	aaa Lys	cgc Arg	tct Ser 175	530
ggg Gly	ttg Leu	gag Glu	acg Thr	gcg Ala 180	acg Thr	tac Tyr	gga Gly	aac Asn	gac Asp 185	gct Ala	gtt Val	ttg Leu	aga Arg	gcg Ala 190	cgt Arg	578
gag Glu	gtt Val	ttg Leu	ttc Phe 195	gag Glu	aag Lys	act Thr	gtt Val	acg Thr 200	ccg Pro	agc Ser	gac Asp	gtc Val	999 Gly 205	aag Lys	ctg Leu	626
aac Asn	cgt Arg	tta Leu 210	gtg Val	ata Ile	ccg Pro	aaa Lys	caa Gln 215	cac His	gcg Ala	gag Glu	aag Lys	cat His 220	ttt Phe	ccg Pro	tta Leu	674
ccg Pro	gcg Ala 225	atg Met	acg Thr	acg Thr	gcg Ala	atg Met 230	ggg ggg	atg Met	aat Asn	ccg Pro	tct Ser 235	ccg Pro	acg Thr	aaa Lys	ggc Gly	722
	ttg Leu															770
tac Tyr	agt Ser	tac Tyr	tgg Trp	aac Asn 260	agc Ser	agt Ser	caa Gln	agt Ser	tac Tyr 265	gtg Val	ttg Leu	acc Thr	aag Lys	ggc Gly 270	tgg Trp	818
agc Ser	cgg Arg	ttc Phe	gtt Val 275	aaa Lys	gag Glu	aag Lys	aat Asn	ctt Leu 280	cga Arg	gcc Ala	ggt Gly	gat Asp	gtg Val 285	gtt Val	tgt Cys	866
ttc Phe	gag Glu	aga Arg 290	tca Ser	acc Thr	gga Gly	cca Pro	gac Asp 295	cgg Arg	caa Gln	ttg Leu	tat Tyr	atc Ile 300	cac His	tgg Trp	aaa Lys	914
gtc	cgg	tct	agt	ccg	gtt	cag	act	gtg		agg age		ttc	gga	gtc	aac	962

Val Arg Ser S	er Pro Val Glr	mbi19 Sequence	Listing.ST25 Leu Phe Gly Val Asn	
305	310		315	
att ttc aat g Ile Phe Asn V 320	tg agt aac gag al Ser Asn Glu 325	aaa cca aac gac Lys Pro Asn Asp 330	gtc gca gta gag tgt 10 Val Ala Val Glu Cys 335	010
gtt ggc aag a Val Gly Lys L	ag aga tot ogg ys Arg Ser Arg 340	gaa gat gat ttg Glu Asp Asp Leu 345	ttt tcg tta ggg tgt 10 Phe Ser Leu Gly Cys 350	58
Ser Lys Lys G	ag gcg att ato ln Ala Ile Ile 55	aac atc ttg tga Asn Ile Leu 360	caaattottt tttttttggtt 11	111
tttttcttca at	ttgtttct ccttt	ttcaa tattttgtat	tgaaatgaca agttgtaaat 11	171
taggacaaga ca	agaaaaaa tgaca	actag acaaaatagt	ttttgtttaa aaaaaaaaaa 12	231
aaaaaaa			12	239
	opsis thaliana			
<400> 58			T) 0 . 01 0 1	
Met Glu Tyr S	er Cys Val Asp 5	Asp Ser Ser Thr 10	Thr Ser Glu Ser Leu 15	
Ser Ile Ser T		Pro Thr Thr Thr 25	Thr Glu Lys Lys Leu 30	
Ser Ser Pro P 35	ro Ala Thr Ser	Met Arg Leu Tyr 40	Arg Met Gly Ser Gly 45	
Gly Ser Ser V	al Val Leu Asp 55	Ser Glu Asn Gly	Val Glu Thr Glu Ser 60	
Arg Lys Leu P 65	ro Ser Ser Lys 70	Tyr Lys Gly Val 75	Val Pro Gln Pro Asn 80	
Gly Arg Trp G	ly Ala Gln Ile 85	Tyr Glu Lys His 90	Gln Arg Val Trp Leu 95	
	sn Glu Glu Glu 00	Glu Ala Ala Ser 105	Ser Tyr Asp Ile Ala 110	
Val Arg Arg P	he Arg Gly Arg	Asp Ala Val Thr 120	Asn Phe Lys Ser Gln 125	
Val Asp Gly A 130	sn Asp Ala Glu 135		Asp Ala His Ser Lys 140	
Ala Glu Ile V 145	al Asp Met Leu 150	Arg Lys His Thr 155	Tyr Ala Asp Glu Phe	
Glu Gln Ser A	rg Arg Lys Phe 165	Val Asn Gly Asp 170	Gly Lys Arg Ser Gly 175	
	la Thr Tyr Gly 80	Asn Asp Ala Val	Leu Arg Ala Arg Glu 190	

### mbi19 Sequence Listing.ST25

Val Leu Phe Glu Lys Thr Val Thr Pro Ser Asp Val Gly Lys Leu Asn Arg Leu Val Ile Pro Lys Gln His Ala Glu Lys His Phe Pro Leu Pro Ala Met Thr Thr Ala Met Gly Met Asn Pro Ser Pro Thr Lys Gly Val Leu Ile Asn Leu Glu Asp Arg Thr Gly Lys Val Trp Arg Phe Arg Tyr 245 250 255Ser Tyr Trp Asn Ser Ser Gln Ser Tyr Val Leu Thr Lys Gly Trp Ser 260 265 270 Arg Phe Val Lys Glu Lys Asn Leu Arg Ala Gly Asp Val Val Cys Phe Glu Arg Ser Thr Gly Pro Asp Arg Gln Leu Tyr Ile His Trp Lys Val Arg Ser Ser Pro Val Gln Thr Val Val Arg Leu Phe Gly Val Asn Ile Phe Asn Val Ser Asn Glu Lys Pro Asn Asp Val Ala Val Glu Cys Val Gly Lys Lys Arg Ser Arg Glu Asp Asp Leu Phe Ser Leu Gly Cys Ser Lys Lys Gln Ala Ile Ile Asn Ile Leu <210> 59 <211> 803 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (35)..(658) <223> G41 <400> 59 55 ctgatcaatg aactcatttt ctgccttttc tgaa atg ttt ggc tcc gat tac gag Met Phe Gly Ser Asp Tyr Glu tct ccg gtt tcc tca ggc ggt gat tac agt ccg aag ctt gcc acg agc Ser Pro Val Ser Ser Gly Gly Asp Tyr Ser Pro Lys Leu Ala Thr Ser 10 15 20 103 tgc ccc aag aaa cca gcg gga agg aag aag ttt cgt gag act cgt cac Cys Pro Lys Lys Pro Ala Gly Arg Lys Lys Phe Arg Glu Thr Arg His 151 cca att tac aga gga gtt cgt caa aga aac tcc ggt aag tgg gtg tgt Pro Ile Tyr Arg Gly Val Arg Gln Arg Asn Ser Gly Lys Trp Val Cys 40 45 . 50 50199 gag ttg aga gag cca aac aag aaa acg agg att tgg ctc ggg act ttc Page 79

Glu	Leu	Arg	Glu		Asn	Lys			Arg		List Trp			Thr	Phe	
<b>~</b> 22	200	cat	~~~	60	~~~	aat	cat	act	65	<b>a</b> aa	gtc	900	acc	70 ata	act	295
Gln	Thr	Ala	Glu 75	Met	Ala	Ala	Arg	Ala 80	His	Asp	Val	Ala	Ala 85	Ile	Ala	
ctc Leu	cgt Arg	ggc Gly 90	aga Arg	tct Ser	gcc Ala	tgt Cys	ct <b>c</b> Leu 95	aat Asn	ttc Phe	gct Ala	gac Asp	tcg Ser 100	gct Ala	tgg Trp	cgg Arg	343
cta Leu	cga Arg 105	atc Ile	ccg Pro	gaa Glu	tca Ser	acc Thr 110	tgt Cys	gcc Ala	aag Lys	gaa Glu	atc Ile 115	caa Gln	aag Lys	gcg Ala	gcg Ala	391
gct Ala 120	gaa Glu	gcc Ala	gcg Ala	ttg Leu	aat Asn 125	ttt Phe	caa Gln	gat Asp	gag Glu	atg Met 130	tgt Cys	cat His	atg Met	acg Thr	acg Thr 135	439
gat Asp	gct Ala	cat His	ggt Gly	ctt Leu 140	gac Asp	atg Met	gag Glu	gag Glu	acc Thr 145	ttg Leu	gtg Val	gag Glu	gct Ala	att Ile 150	tat Tyr	487
acg Thr	ccg Pro	gaa Glu	cag Gln 155	agc Ser	caa Gln	gat Asp	gcg Ala	ttt Phe 160	tat Tyr	atg Met	gat Asp	gaa Glu	gag Glu 165	gcg Ala	atg Met	535
ttg Leu	Gly ggg	atg Met 170	tct Ser	agt Ser	ttg Leu	ttg Leu	gat Asp 175	aac Asn	atg Met	gcc Ala	gaa Glu	999 Gly 180	atg Met	ctt Leu	tta Leu	583
											gat Asp 195					631
		gtg Val						taa	aatt	.cgat	tt t	tatt	tcca	t		678
ttt	ggta	att a	tago	tttt	t at	acat	ttga	tcc	ttt	tta	gaat	ggat:	ct t	cttc	ttttt	738
ttgg	ttgt	ga g	jaaac	gaat	g ta	aatg	gtaa	aag	ttgt	tgt	caaa	tgca	aa t	gttt	ttgag	798
tgca	ıg															803
<210 <211 <212 <213	> 2 !> E	50 207 PRT Arabi	.dops	is t	hali	.ana										
<400	)> 6	0														
Met 1	Phe	Gly	Ser	Asp 5	Tyr	Glu	Ser	Pro	Val 10	Ser	Ser	Gly	Gly	Asp 15	Tyr	
Ser	Pro	Lys	Leu 20	Ala	Thr	Ser	Сув	Pro 25	Lys	Lys	Pro	Ala	Gly 30	Arg	Lys	
Lys	Phe	Arg 35	Glu	Thr	Arg	His	Pro 40	Ile	Tyr	Arg	Gly	Val 45	Arg	Gln	Arg	
Asn	Ser 50	Gly	Lys	Trp	Val	Cys 55	Glu	Leu	Arg	Glu	Pro 60	Asn	Lys	Lys	Thr	
Arg 65	Ile	Trp	Leu	Gly	Thr 70	Phe	Gln	Thr	Ala	Glu 75	Met	Ala	Ala	Arg	Ala 80	
His	Asp	Val	Ala	Ala 85	Ile	Ala	Leu	Arg	90		Ser	Ala	Сув	Leu 95	Asn	
									Pē	ige 8	5 U					

#### mbil9 Sequence Listing.ST25

Phe Ala Asp Ser Ala Trp Arg Leu Arg Ile Pro Glu Ser Thr Cys Ala Lys Glu Ile Gln Lys Ala Ala Ala Glu Ala Ala Leu Asn Phe Gln Asp Glu Met Cys His Met Thr Thr Asp Ala His Gly Leu Asp Met Glu Glu Thr Leu Val Glu Ala Ile Tyr Thr Pro Glu Gln Ser Gln Asp Ala Phe Tyr Met Asp Glu Glu Ala Met Leu Gly Met Ser Ser Leu Leu Asp Asn Met Ala Glu Gly Met Leu Leu Pro Ser Pro Ser Val Gln Trp Asn Tyr Asn Phe Asp Val Glu Gly Asp Asp Asp Val Ser Leu Trp Ser Tyr <210> 61 <211> 929 <212> DNA <213> Arabidopsis thaliana <220> <221> CDS <222> (164)..(805) <223> G40 <400> 61 cttgaaaaag aatctacctg aaaagaaaaa aaagagagag agatataaat agctttacca 60 agacagatat actatctttt attaatccaa aaagactgag aactctagta actacgtact 120 175 acttaaacct tatccagttt cttgaaacag agtactctga tca atg aac tca ttt Met Asn Ser Phe tca gct ttt tct gaa atg ttt ggc tcc gat tac gag cct caa ggc gga Ser Ala Phe Ser Glu Met Phe Gly Ser Asp Tyr Glu Pro Gln Gly Gly 223 10 gat tat tgt ccg acg ttg gcc acg agt tgt ccg aag aaa ccg gcg ggc Asp Tyr Cys Pro Thr Leu Ala Thr Ser Cys Pro Lys Lys Pro Ala Gly 271 cgt aag aag ttt cgt gag act cgt cac cca att tac aga gga gtt cgt Arg Lys Lys Phe Arg Glu Thr Arg His Pro Ile Tyr Arg Gly Val Arg 319 caa aga aac tcc ggt aag tgg gtt tct gaa gtg aga gag cca aac aag Gln Arg Asn Ser Gly Lys Trp Val Ser Glu Val Arg Glu Pro Asn Lys 55 60367 aaa acc agg att tgg ctc ggg act ttc caa acc gct gag atg gca gct Lys Thr Arg Ile Trp Leu Gly Thr Phe Gln Thr Ala Glu Met Ala Ala 70 75 80415 cgt gct cac gac gtc gct gca tta gcc ctc cgt ggc cga tca gca tgt Arg Ala His Asp Val Ala Ala Leu Ala Leu Arg Gly Arg Ser Ala Cys 463 ctc aac ttc gct gac tcg gct tgg cgg cta cga atc ccg gag tca aca Page 81

Leu Asr	Phe	Ala	Asp 105	Ser	Ala	mbi Trp	19 S Arg	eque Leu 110	ncė Arg	List Ile	ing. Pro	ST25 Glu	Ser 115	Thr	
tgc gcc Cys Ala	aag Lys	gat Asp 120	atc Ile	caa Gln	aaa Lys	gcg Ala	gct Ala 125	gct Ala	gaa Glu	gcg Ala	gcg Ala	ttg Leu 130	gct Ala	ttt Phe	559
caa gat Gln Asp	gag Glu 135	acg Thr	tgt Cys	gat Asp	acg Thr	acg Thr 140	acc Thr	acg Thr	aat Asn	cat His	ggc Gly 145	ctg Leu	gac Asp	atg Met	607
gag gag Glu Glu 150	Thr	atg Met	gtg Val	gaa Glu	gct Ala 155	att Ile	tat Tyr	aca Thr	ccg Pro	gaa Glu 160	cag Gln	agc Ser	gaa Glu	ggt Gly	655
gcg ttt Ala Phe 165	tat Tyr	atg Met	gat Asp	gag Glu 170	gag Glu	aca Thr	atg Met	ttt Phe	999 Gly 175	atg Met	ccg Pro	act Thr	ttg Leu	ttg Leu 180	703
gat aat Asp Asr	atg Met	gct Ala	gaa Glu 185	ggc Gly	atg Met	ctt Leu	tta Leu	ccg Pro 190	ccg Pro	ccg Pro	tct Ser	gtt Val	caa Gln 195	tgg Trp	751
aat cat Asn His	aat Asn	tat Tyr 200	gac Asp	ggc Gly	gaa Glu	gga Gly	gat Asp 205	ggt Gly	gac Asp	gtg Val	tcg Ser	ctt Leu 210	tgg Trp	agt Ser	799
tac taa Tyr	tati	cgat	ag t	cgtt	tcca	at t	ttgt	acta	a tag	gttt	gaaa	atat	teta	agt	855
tcctttt	ttt a	agaat	ggtt	c ct	tcat	ttta	a ttt	tati	tta	ttgi	tgta	nga a	acga	agtgga	915
aaataat	tca a	atac													929
<210><211><211><212><213>	62 213 PRT Arab:	idops	sis t	:hali	iana										
<211> <212>	213 PRT Arab:	idops	sis t	:hali	iana										
<211> <212> <213>	213 PRT Arab:	_				Ser	Glu	Met 10	Phe	Gly	Ser	Asp	Туг 15	Glu	
<211><212><213><400> '	213 PRT Arab: 62 Ser	Phe	Ser 5	Ala	Phe			10					15		
<211> <212> <213> <400> '	213 PRT Arab 62 Ser	Phe Gly 20	Ser 5 Asp	Ala Tyr	Phe Cys	Pro	Thr 25	10 Leu	Ala	Thr	Ser	Сув 30	15 Pro	Lys	
<211> <212> <213> <400>  Met Asm  1  Pro Glm	213 PRT Arab: 62 Ser Gly Ala 35	Phe Gly 20	Ser 5 Asp Arg	Ala Tyr Lys	Phe Cys Lys	Pro Phe 40	Thr 25 Arg	10 Leu Glu	Ala Thr	Thr	Ser His	Cys 30 Pro	Pro	Lys Tyr	
<211> <212> <213> <400>  Met Asm  1  Pro Glm  Lys Pro	213 PRT Arab: 62 Ser Gly Ala 35	Phe Gly 20 Gly	Ser 5 Asp Arg	Ala Tyr Lys Arg	Phe Cys Lys Asn 55	Pro Phe 40	Thr 25 Arg	Leu Glu Lys	Ala Thr	Thr Arg Val	Ser His 45 Ser	Cys 30 Pro Glu	Pro Ile Val	Lys Tyr Arg	
<211> <212> <213> <400>  Met Asm  1  Pro Glm  Lys Pro  Arg Gly 50  Glu Pro	213 PRT Arab: 62 Ser Gly Ala 35 Val	Phe Gly 20 Gly Arg	Ser 5 Asp Arg Gln	Ala Tyr Lys Arg	Phe Cys Lys Asn 55	Pro Phe 40 Ser	Thr 25 Arg Gly Trp	Leu Glu Lys Leu	Ala Thr Trp Gly	Thr Arg Val 60	Ser His 45 Ser	Cys 30 Pro Glu	Pro Ile Val	Lys Tyr Arg Ala 80	
<211> <212> <213> <400>  Met Asm  1  Pro Glm  Lys Pro  Arg Gly  50  Glu Pro  65	213 PRT Arab 62 Ser Gly Ala 35 Val Asn	Phe Gly 20 Gly Arg Lys	Ser 5 Asp Arg Gln Lys Arg 85	Tyr Lys Arg Thr 70	Phe Cys Lys Asn 55 Arg	Pro Phe 40 Ser Ile	Thr 25 Arg Gly Trp	Leu  Glu  Lys  Leu  Ala  90	Ala Thr Trp Gly 75 Ala	Thr Arg Val 60 Thr	Ser His 45 Ser Phe	Cys 30 Pro Glu Gln Leu	Pro Ile Val Thr Arg	Lys Tyr Arg Ala 80 Gly	

Ala	Leu 130	Ala	Phe	Gln	Asp	Glu 135	mbi Thr	19 S Cys	Asp Ague	nce Thr	List Thr 140	ing.: Thr	ST25 Thr	Asn	His	
Gly 145	Leu	Авр	Met	Glu	Glu 150	Thr	Met	Val	Glu	Ala 155	Ile	Tyr	Thr	Pro	Glu 160	
Gln	Ser	Glu	Gly	Ala 165	Phe	туг	Met	Авр	Glu 170	Glu	Thr	Met	Phe	Gly 175	Met	
Pro	Thr	Leu	Leu 180	Asp	Asn	Met	Ala	Glu 185	Gly	Met	Leu	Leu	Pro 190	Pro	Pro	
Ser	Val	Gln 195	Trp	Asn	His	Asn	Tyr 200	Asp	Gly	Glu	Gly	Asp 205	Gly	Asp	Val	
Ser	Leu 210	Trp	Ser	Tyr												
<210 <211 <212 <213	> ! > I	53 908 DNA Arab:	idops	sis (	hali	iana										
<220 <221 <222 <223	> ( >	CDS (119) 342	(	769)												
<400 cctg		3 ag a	aacag	gaaag	ga ga	agaga	aaact	t att	atti	cag	caaa	accat	ac o	caaca	aaaaa	60
										_						
gaca	gaga	atc (	ttta	agtta	ac ct											118
gaca atg Met 1	aac	tca	ttt	·tct	qct	tato	cagi	t tto	cttga	aaac ttt	agag ggc	gtact	ct 1	tac	atca gag	118 166
atg Met	aac Asn	tca Ser	ttt Phe	tct Ser 5	gct Ala	ttato ttt Phe	tct Ser	gaa Glu tat	atg Met 10	ttt Phe	agag ggc Gly acg	tcc Ser	gat Asp	tac Tyr 15	gag Glu agc	
atg Met 1	aac Asn tcg Ser	tca Ser gtt Val	ttt Phe tcc Ser 20	tct Ser 5 tca Ser	gct Ala ggc Gly	ttt The ggt Gly	tct Ser gat Asp	gaa Glu tat Tyr 25	atg Met 10 att Ile	ttt Phe ccg Pro	agag ggc Gly acg Thr	tcc Ser Ctt Leu	gat Asp gcg Ala 30	tac Tyr 15 agc Ser	gag Glu agc Ser	166
atg Met 1 tct Ser tgc Cys	aac Asn tcg Ser ccc Pro	tca Ser gtt Val aag Lys 35	ttt Phe tcc Ser 20 aaa Lys	tca ser ccg Pro	gct Ala ggc Gly gcg Ala	ttato ttt Phe ggt Gly ggt Gly	tct Ser gat Asp cgt Arg	gaa Glu tat Tyr 25 aag Lys	atg Met 10 att Ile aag Lys	ttt Phe ccg Pro ttt Phe	agag ggc Gly acg Thr cgt Arg	tcc Ser ctt Leu gag Glu 45	gat Asp gcg Ala 30 act Thr	tac Tyr 15 agc Ser cgt Arg	gag Glu agc Ser cac His	166 214
atg Met 1 tct Ser tgc Cys	aac Asn tcg Ser ccc Pro ata Ile	tca Ser gtt Val aag Lys 35 tac Tyr	ttt Phe tcc Ser 20 aaa Lys aga Arg	tct ser 5 tca ser ccg Pro gga Gly	gct Ala ggc Gly gcg Ala gtt Val	ttato Phe ggt Gly ggt Gly cgt Arg 55	tct Ser gat Asp cgt Arg 40 cgg Arg	gaa Glu tat Tyr 25 aag Lys aga Arg	atg Met 10 att Ile aag Lys aac Asn	ttt Phe ccg Pro ttt Phe tcc Ser	agag ggc Gly acg Thr cgt Arg ggt Gly 60	tcc Ser ctt Leu gag Glu 45 aag Lys	gat Asp gcg Ala 30 act Thr tgg Trp	tac Tyr 15 agc Ser cgt Arg gtt Val	gag Glu agc Ser cac His tgt Cys	166 214 262
atg Met 1 tct ser tgc Cys cca Pro	aac Asn tcg Ser ccc Pro ata Ile 50 gtt	tca Ser gtt Val aag Lys 35 tac Tyr aga Arg	tttt Phe tcc Ser 20 aaa Lys aga Arg gaa Glu	tct Ser 5 tca Ser ccg Pro gga Gly cca Pro	gct Ala ggc Gly gcg Ala gtt Val aac Asn 70	ttatc ttt Phe ggt Gly ggt Gly cgt Arg 55 aag Lys	tct Ser gat Asp cgt Arg cgg Arg aaa Lys	gaa Glu tat Tyr 25 aag Lys aga Arg	atg Met 10 att Ile aag Lys aac Asn agg Arg	ttt Phe ccg Pro ttt Phe tcc Ser att Ile 75	agag ggc Gly acg Thr cgt Arg ggt Gly 60 tgg	tcc Ser Ctt Leu gag Glu 45 aag Lys Ctc Leu	gat Asp gcg Ala 30 act Thr tgg gray gct gct	tac Tyr 15 agc Ser cgt Arg gtt Val aca Thr	gag Glu agc Ser cac His tgt Cys ttt Phe 80	166 214 262 310
atg Met 1 tct Ser tgc Cys cca Pro	aaccAsn tcgSer cccCPro ata ile 50 gtt Val accThr	tca Ser gtt Val aag Lys 35 tac Tyr aga Arg	ttt Phe tcc Ser 20 aaa Lys aga Arg gaa Glu	tct 5 tca Ser ccg Pro gga Gly cca Pro atg Met 85	gct Ala ggc Gly gcg Ala gtt Val aacc Asn 70 gca Ala	ttatc ttt Phe ggt Gly ggt Arg 55 aag Ly8 gct Ala	tct Ser gat Asp cgt Arg Arg Arg ctc	gaa Glu tat Tyr 25 aag Lys aga Arg aca Thr	atg Met 10 att Ile aag Lys aac Asn agg Arg	tttt Phe ccg Pro tttt Phe tcc Ser att Ile 75 gac Asp gct	agag ggc Gly acg Thr cgt Arg ggt Gly 60 tgg Trp	tcc Ser ctt Leu gag Glu 45 aag Lys ctc Leu	gat Asp gcg Ala 30 act Thr tgg Trp gga Gly gct Ala gct	tac Tyr 15 agc Ser cgt Arg gtt Val aca Thr tta Leu 95 tgg	gag Glu agc Ser cac His tgt Cys ttt Phe 80 gcc Ala	166 214 262 310
atg Met 1 tct Ser tgc Cys cca Pro gag Glu 65 caa Gln	aacc Asn tcg Ser ccc Pro ata Ile 50 gtt Val acc Thr cgt Arg	tca Ser gtt Val aag Lys 35 tac Tyr aga Arg gct Ala	ttt Phe tcc Ser 20 aaa Lys aga Arg Glu gag Glu cga Arg 100	tct Ser 5 tca Ser ccg Pro gga Gly cca Pro atg Met 85 tca Ser	gct Ala ggc Gly gcg Ala gtt Val aacc Asn 70 gca Ala gcc Ala	ttatc ttt Phe ggt Gly ggt Arg tys gct Ala tgt Cys act	tct Ser gat Asp cgt Arg Arg Arg cgz Arg tys Cga Arg ttc Leu tqc	gaa Glu tat Tyr 25 aag Lys aga Arg aca Thr gct Ala	atg Met 10 att Ile aag Lys aac Asn agg Arg cac His 90 ttc Phe	tttt Phe ccg Pro tttt Phe tcc Ser attt 175 gac Asp gct Ala	agag ggc Gly acg Thr cgt Arg ggt Gly 60 tgg Trp gtt Val	tcc Ser ctt Leu gag Glu 45 aag Lys ctc Leu gcc Ala tcg Ser	gct Asp gcg Ala 30 act Thr tgg Gly gct Ala gct Ala ang aag	tac Tyr 15 agc Ser cgt Arg gtt Val aca Thr tta Leu 95 tgg Trp gcg	gag Glu agc Ser cac His tgt Cys ttt Phe 80 gcc Ala aga Arg	166 214 262 310 358 406

Page 83

## mbil9 Sequence Listing.ST25

								•			5				
gat c Asp H 145	at ggo lis Gly	ttc Phe	gac Asp	atg Met 150	gag Glu	gag Glu	acg Thr	ttg Leu	gtg Val 155	gag Glu	gct Ala	att Ile	tac Tyr	acg Thr 160	598
gcg g Ala G	aa cag lu Glr	agc Ser	gaa Glu 165	aat Asn	gcg Ala	ttt Phe	tat Tyr	atg Met 170	cac His	gat Asp	gag Glu	gcg Ala	atg Met 175	ttt Phe	646
gag a Glu M	tg ccg let Pro	agt Ser 180	ttg Leu	ttg Leu	gct Ala	aat Asn	atg Met 185	gca Ala	gaa Glu	999 Gly	atg Met	ctt Leu 190	ttg Leu	ccg Pro	694
ctt c Leu P	cg tcc ro Ser 195	val	cag Gln	tgg Trp	aat Asn	cat His 200	aat Asn	cat His	gaa Glu	gtc Val	gac Asp 205	ggc Gly	gat Asp	gat Asp	742
Asp A	ac gta sp Val	tcg Ser	tta Leu	tgg Trp	agt Ser 215	tat Tyr	taa	aaci	tcaga	att a	attai	tttc	ca		789
tttt	agtac	gata	cttt	t at	tttta	atta	t ta	tttt	taga	tcc	tttt	tta 🤉	gaate	ggaatc	849
tncat	tatgt	ttgt	aaaa	ct ga	agaaa	acga	g tg	taaa	ttaa	att	gatto	cag	tttca	agtat	908
<210><211><211><212><213>	216 PRT Arab	oidopa	sis t	hali	iana										
		Dha	Com	212	Dho	Cor	C2	Wot	Dho	Cl.	cor	ممر	Tur	Gly	
Met A	sn Ser	Pne	5	Ala	Pne .	ser	GIU	10	Pne	GIY	Ser	мар	15	GIU	
Ser S	er Val	Ser 20	Ser	Gly	Gly	Asp	Tyr 25	Ile	Pro	Thr	Leu	Ala 30	Ser	Ser	
Cys P	ro Lys 35	Lys	Pro	Ala	Gly	Arg 40	Lys	Lys	Phe	Arg	Glu 45	Thr	Arg	His	
	le Tyr O	Arg	Gly	Val	Arg 55	Arg	Arg	Aen	Ser	Gly 60	Lys	Trp	Val	Cys	
Glu V	al Arg	Glu	Pro	Asn 70	Lys	Lys	Thr	Arg	Ile 75	Trp	Leu	Gly	Thr	Phe 80	
Gln T	hr Ala		Met 85		Ala	Arg	Ala	His 90		Val	Ala	Ala	Leu 95	Ala	
Leu A	rg Gly	Arg 100	Ser	Ala	Сув	Leu	Asn 105	Phe	Ala	Asp	Ser	Ala 110	Trp	Arg	
Leu A	rg Ile 115		Glu	Ser	Thr	Cys 120	Ala	Lys	Asp	Ile	Gln 125	Lys	Ala	Ala	
	lu Ala 30	Ala	Leu	Ala	Phe 135	Gln	Asp	Glu	Met	Cys 140	Asp	Ala	Thr	Thr	
Asp H.	is Gly	Phe	Asp	Met 150	Glu	Glu	Thr	Leu	Val 155	Glu	Ala	Ile	Tyr	Thr 160	
Ala G	lu Gln	Ser	Glu 165	Asn	Ala	Phe	Tyr	Met 170	His	Asp	Glu	Ala	Met 175	Phe	

Page 84

### mbi19 Sequence Listing.ST25

Glu Met Pro Ser Leu Leu Ala Asn Met Ala Glu Gly Met Leu Leu Pro 180 185 190

Leu Pro Ser Val Gln Trp Asn His Asn His Glu Val Asp Gly Asp Asp

Asp Asp Val Ser Leu Trp Ser Tyr

<210> 65
<211> 1407
<212> DNA
<213> Arabidopsis thaliana
<220>
<221> CDS
<222> (191)..(1351)
<223> G1127

<400> 65 60 gacagactct ctctgtatgt gtgcgagaag cgagaagcga gagagagaga gagagagttg ttagctcaca cgctttctct attttctcgg aattcacaaa acagaaagtt tcatccttta 120 cgagaattaa gccgaaagaa acaatctttg agtttgattt cttcttcctt ccttctctct 180 ctctgctcta atg gat tcc aga gac atc cca ccg tca cat aac cag ctt Met Asp Ser Arg Asp Ile Pro Pro Ser His Asn Gln Leu 229 caa cca cca ccg gga atg tta atg tct cat tac cgt aac cct aac gcc Gln Pro Pro Pro Gly Met Leu Met Ser His Tyr Arg Asn Pro Asn Ala 277 gcc gct tca cca tta atg gtt ccc act tcc aca tct caa ccg att caa Ala Ala Ser Pro Leu Met Val Pro Thr Ser Thr Ser Gln Pro Ile Gln 325 40 373 cac cct cgt ctt cct ttt ggc aat caa caa caa tct caa acg ttt cat His Pro Arg Leu Pro Phe Gly Asn Gln Gln Ser Gln Thr Phe His cag cag caa caa caa caa atg gat cag aag act ctt gaa tct ctt gga Gln Gln Gln Gln Gln Met Asp Gln Lys Thr Leu Glu Ser Leu Gly 421 ttt ggt gat gga tca cct tct tct caa ccg atg cga ttc ggg atc gat Phe Gly Asp Gly Ser Pro Ser Ser Gln Pro Met Arg Phe Gly Ile Asp 469 gat cag aat cag caa ctg caa gtg aag aag cga gga agg ccg aga Asp Gln Asn Gln Gln Leu Gln Val Lys Lys Lys Arg Gly Arg Pro Arg 517 aag tat act cct gat ggt agc att gct tta ggt tta gct cct acg tct Lys Tyr Thr Pro Asp Gly Ser Ile Ala Leu Gly Leu Ala Pro Thr Ser 565 cct ctt ctc tct gca gct tct aat tct tac ggt gag ggt ggt ggt gga Pro Leu Leu Ser Ala Ala Ser Asn Ser Tyr Gly Glu Gly Gly Val Gly 613 gat agt ggt gga aat gga aac tot gtt gat coa cot gtt aaa cgt aac Asp Ser Gly Gly Asn Gly Asn Ser Val Asp Pro Pro Val Lys Arg Asn 661 150

Page 85

709

aga gga agg cct cct ggt tct agt aag aaa cag ctt gat gct tta gga Arg Gly Arg Pro Pro Gly Ser Ser Lys Lys Gln Leu Asp Ala Leu Gly 165 170

## mbil9 Sequence Listing.ST25

									-			-				
gga Gly	act Thr 175	tca Ser	gga Gly	gtt Val	999 Gly	ttt Phe 180	aca Thr	cct Pro	cat His	gtc Val	att Ile 185	gaa Glu	gtg Val	aac Asn	aca Thr	757
gga Gly 190	gag Glu	gac Asp	ata Ile	gcg Ala	tca Ser 195	aag Lys	gtg Val	atg Met	gct Ala	ttt Phe 200	tcg Ser	gat Asp	caa Gln	999 Gly	tca Ser 205	805
aga Arg	aca Thr	att Ile	tgt Cys	att Ile 210	ctc Leu	tct Ser	gca Ala	agt Ser	ggt Gly 215	gca Ala	gtt Val	tct Ser	aga Arg	gtg Val 220	atg Met	853
ctt Leu	cgt Arg	caa Gln	gct Ala 225	tct Ser	cat His	tct Ser	agt Ser	gga Gly 230	atc Ile	gtt Val	act Thr	tat Tyr	gag Glu 235	gga Gly	cga Arg	901
ttt Phe	gag Glu	atc Ile 240	att Ile	act Thr	ctc Leu	tca Ser	ggc Gly 245	tca Ser	gtc Val	ttg Leu	aat Asn	tat Tyr 250	gag Glu	gta Val	aat Asn	949
ggt Gly	tcc Ser 255	acc Thr	aac Asn	aga Arg	agt Ser	ggt Gly 260	aac Asn	ttg Leu	agt Ser	gtg Val	gct Ala 265	ttg Leu	gct Ala	gga Gly	cct Pro	997
gat Asp 270	ggc Gly	ggc Gly	atc Ile	gta Val	ggt Gly 275	ggc Gly	agt Ser	gta Val	gtt Val	ggt Gly 280	aat Asn	cta Leu	gta Val	gct Ala	gca Ala 285	1045
aca Thr	caa Gln	gtc Val	cag Gln	gtg Val 290	ata Ile	gtg Val	gga Gly	agc Ser	ttt Phe 295	gtt Val	gca Ala	gaa Glu	gca Ala	aag Lys 300	aaa Lys	1093
										999 Gly						1141
gct Ala	tca Ser	gcg Ala 320	ccg Pro	gct Ala	aac Asn	atg Met	ttg Leu 325	aac Asn	ttt Phe	gga Gly	tca Ser	gtc Val 330	tct Ser	caa Gln	gga Gly	1189
cca Pro	tcg Ser 335	agc Ser	gag Glu	tca Ser	tca Ser	gaa Glu 340	gag Glu	aat Asn	gag Glu	agc Ser	ggt Gly 345	tct Ser	cct Pro	gca Ala	atg Met	1237
										gct Ala 360						1285
caa Gln	caa Gln	cct Pro	ctt Leu	cat His 370	cct Pro	cat His	cag Gln	atg Met	caa Gln 375	atg Met	tac Tyr	caa Gln	cat His	ctt Leu 380	tgg Trp	1333
	aat Asn				taa	aatg	aago	gg a	aatt	aatt	t gt	ttco	gttt	:		1381
tggt	tacg	ıgt t	atgg	tttg	ga tt	tett	:									1407

<210> 66 <211> 386 <212> PRT <213> Arabidopsis thaliana

<400> 66

Met Asp Ser Arg Asp Ile Pro Pro Ser His Asn Gln Leu Gln Pro Pro 1 0 15

Pro Gly Met Leu Met Ser His Tyr Arg Asn Pro Asn Ala Ala Ser 20 25 30

mbil9 Sequence Listing.ST25
Pro Leu Met Val Pro Thr Ser Thr Ser Gln Pro Ile Gln His Pro Arg
35 40 45

Leu Pro Phe Gly Asn Gln Gln Gln Ser Gln Thr Phe His Gln Gln Gln 50 55 60

Gln Gln Gln Met Asp Gln Lys Thr Leu Glu Ser Leu Gly Phe Gly Asp 65 70 75

Gly Ser Pro Ser Ser Gln Pro Met Arg Phe Gly Ile Asp Asp Gln Asn 85 90 95

Gln Gln Leu Gln Val Lys Lys Lys Arg Gly Arg Pro Arg Lys Tyr Thr

Pro Asp Gly Ser Ile Ala Leu Gly Leu Ala Pro Thr Ser Pro Leu Leu 115 120 125

Ser Ala Ala Ser Asn Ser Tyr Gly Glu Gly Gly Val Gly Asp Ser Gly 130 140

Gly Asn Gly Asn Ser Val Asp Pro Pro Val Lys Arg Asn Arg Gly Arg 145 150 155 160

Pro Pro Gly Ser Ser Lys Lys Gln Leu Asp Ala Leu Gly Gly Thr Ser 165 170 175

Gly Val Gly Phe Thr Pro His Val Ile Glu Val Asn Thr Gly Glu Asp 180 185 190

The Ala Ser Lys Val Met Ala Phe Ser Asp Gln Gly Ser Arg Thr Ile 195 200 205

Cys Ile Leu Ser Ala Ser Gly Ala Val Ser Arg Val Met Leu Arg Gln 210 215 220

Ala Ser His Ser Ser Gly Ile Val Thr Tyr Glu Gly Arg Phe Glu Ile 225 230 240

Ile Thr Leu Ser Gly Ser Val Leu Asn Tyr Glu Val Asn Gly Ser Thr 245  $\phantom{000}250$   $\phantom{000}255$ 

Asn Arg Ser Gly Asn Leu Ser Val Ala Leu Ala Gly Pro Asp Gly Gly 260 265 270

Ile Val Gly Gly Ser Val Val Gly Asn Leu Val Ala Ala Thr Gln Val 275 280 285

Gln Val Ile Val Gly Ser Phe Val Ala Glu Ala Lys Lys Pro Lys Gln 290 295 300

Ser Ser Val Asn Ile Ala Arg Gly Gln Asn Pro Glu Pro Ala Ser Ala 305 310 320

Pro Ala Asn Met Leu Asn Phe Gly Ser Val Ser Gln Gly Pro Ser Ser 325 330 335

mbi19 Sequence Listing.ST25

Glu Ser Ser Glu Glu Asn Glu Ser Gly Ser Pro Ala Met His Arg Asp

Asn Asn Gly Ile Tyr Gly Ala Gln Gln Gln Gln Gln Gln Pro

Leu His Pro His Gln Met Gln Met Tyr Gln His Leu Trp Ser Asn His

Gly Gln

<210> 67 <211> 1020 <212> DNA

<213> Arabidopsis thaliana

<221> CDS <222> (1)..(1020) <223> G2657

<400> 67 atg gat cca gtt caa tct cat gga tca caa agc tct ctt cct cct cct Met Asp Pro Val Gln Ser His Gly Ser Gln Ser Ser Leu Pro Pro 4 B 96 144 40 cat cag caa cca caa aga aac ctt gat caa gat cac gag cag caa gga His Gln Gln Pro Gln Arg Asn Leu Asp Gln Asp His Glu Gln Gln Gly 50 60 ggg tca ata ttg aat aga tct atc aag atg gat cgc gaa gag aca agc Gly Ser Ile Leu Asn Arg Ser Ile Lys Met Asp Arg Glu Glu Thr Ser 240 gat aac atg gac aac atc gct aat acc aac agc ggt agc gaa ggt aaa Asp Asn Met Asp Asn Ile Ala Asn Thr Asn Ser Gly Ser Glu Gly Lys 288 gag atg agt tta cac gga gga gga gga agc ggt ggt gga gga agt Glu Met Ser Leu His Gly Gly Glu Gly Gly Ser Gly Gly Gly Gly Ser 336 gga gaa cag atg aca aga agg cca aga gga aga cca gca gga tcc aag Gly Glu Gln Met Thr Arg Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys 384 aac aaa cct aaa gct cca ata atc ata aca aga gac agc gca aac gcg Asn Lys Pro Lys Ala Pro Ile Ile Ile Thr Arg Asp Ser Ala Asn Ala 432 480 ctt cga act cac gtc atg gag ata gga gac gga tgt gac ata gtt gac Leu Arg Thr His Val Met Glu Ile Gly Asp Gly Cys Asp Ile Val Asp tgt atg gct acg ttc gct aga cgc cgc caa aga ggc gtt tgc gtt atg Cys Met Ala Thr Phe Ala Arg Arg Gln Arg Gly Val Cys Val Met 528

Page 88

576

agc ggt aca gga agc gtt act aac gtc act ata cgt cag cct gga tcg Ser Gly Thr Gly Ser Val Thr Asn Val Thr Ile Arg Gln Pro Gly Ser

PCT/US00/31414 WO 01/35725

			180				mbi	19 S 185	eque	nce :	List	ing.	ST25 190			
cca Pro	cct Pro	ggc Gly 195	tcg Ser	gtg Val	gtt Val	agc Ser	ctt Leu 200	cac His	ggc Gly	cgg Arg	ttt Phe	gaa Glu 205	atc Ile	ctc Leu	tct Ser	624
ctt Leu	tcg Ser 210	gga Gly	tct Ser	ttc Phe	ttg Leu	cct Pro 215	ccg Pro	cct Pro	gcg Ala	ccg Pro	cct Pro 220	gca Ala	gcc Ala	acc Thr	gga Gly	672
cta Leu 225	agc Ser	gtt Val	tac Tyr	cta Leu	gcc Ala 230	gga Gly	gga Gly	caa Gln	ggg Gly	cag Gln 235	gtc Val	gtt Val	gga Gly	ggt Gly	agt Ser 240	720
gtg Val	gtg Val	gga Gly	cct Pro	ttg Leu 245	ttg Leu	tgt Cys	tcg Ser	ggt Gly	cct Pro 250	gtg Val	gtg Val	gtt Val	atg Met	gcg Ala 255	gct Ala	768
tct Ser	ttt Phe	agc Ser	aat Asn 260	gcg Ala	gcg Ala	tac Tyr	gaa Glu	agg Arg 265	ctg Leu	cct Pro	ttg Leu	gaa Glu	gaa Glu 270	gat Asp	gag Glu	816
atg Met	cag Gln	acg Thr 275	cca Pro	gtt Val	caa Gln	gga Gly	ggc Gly 280	ggt Gly	gga Gly	gga Gly	gga Gly	gga Gly 285	ggt Gly	ggt Gly	ggt Gly	864
gga Gly	atg Met 290	gga Gly	tct Ser	ccc Pro	ccg Pro	atg Met 295	atg Met	gga Gly	cag Gln	caa Gln	caa Gln 300	gct Ala	atg Met	gca Ala	gct Ala	912
atg Met 305	gcg Ala	gcg Ala	gct Ala	caa Gln	gga Gly 310	cta Leu	cca Pro	ccg Pro	aat Asn	ctt Leu 315	ctt Leu	ggt Gly	tcg Ser	gtt Val	cag Gln 320	960
ttg Leu	cca Pro	ccg Pro	cca Pro	caa Gln 325	cag Gln	aat Asn	gat Asp	cag Gln	cag Gln 330	tat Tyr	tgg Trp	tct Ser	acg Thr	ggt Gly 335	cgg Arg	1008
	ccg Pro		tga													1020

<210> 68

<211> 339
<212> PRT
<213> Arabidopsis thaliana

<400> 68

Met Asp Pro Val Gln Ser His Gly Ser Gln Ser Ser Leu Pro Pro Pro 1 5 15

Phe His Ala Arg Asp Phe Gln Leu His Leu Gln Gln Gln Gln Gln His 20 25 30

Gln Gln Gln His Gln Gln Gln Gln Gln Gln Gln Phe Phe Leu His His 35 40 45

His Gln Gln Pro Gln Arg Asn Leu Asp Gln Asp His Glu Gln Gln Gly 50 60

Gly Ser Ile Leu Asn Arg Ser Ile Lys Met Asp Arg Glu Glu Thr Ser 65 70 . 75 80

Asp Asn Met Asp Asn Ile Ala Asn Thr Asn Ser Gly Ser Glu Gly Lys 85 90 95

Glu Met Ser Leu His Gly Gly Glu Gly Gly Ser Gly Gly Gly Ser Page 89

PCT/US00/31414 WO 01/35725

mbi19 Sequence Listing.ST25

Gly Glu Gln Met Thr Arg Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys 115 120 125

100

Asn Lys Pro Lys Ala Pro Ile Ile Ile Thr Arg Asp Ser Ala Asn Ala

Leu Arg Thr His Val Met Glu Ile Gly Asp Gly Cys Asp Ile Val Asp 145 150 160

Cys Met Ala Thr Phe Ala Arg Arg Arg Gln Arg Gly Val Cys Val Met 165 170 175

Ser Gly Thr Gly Ser Val Thr Asn Val Thr Ile Arg Gln Pro Gly Ser 180 185 190

Pro Pro Gly Ser Val Val Ser Leu His Gly Arg Phe Glu Ile Leu Ser 195 200 205

Leu Ser Gly Ser Phe Leu Pro Pro Pro Ala Pro Pro Ala Ala Thr Gly 210 215 220

Leu Ser Val Tyr Leu Ala Gly Gly Gln Gly Gln Val Val Gly Gly Ser

Val Val Gly Pro Leu Leu Cys Ser Gly Pro Val Val Wat Ala Ala 245 250 255

Ser Phe Ser Asn Ala Ala Tyr Glu Arg Leu Pro Leu Glu Glu Asp Glu 260 265 270

Gly Met Gly Ser Pro Pro Met Met Gly Gln Gln Ala Met Ala Ala

Met Ala Ala Gln Gly Leu Pro Pro Asn Leu Leu Gly Ser Val Gln 305 310 315 320

Leu Pro Pro Pro Gln Gln Asn Asp Gln Gln Tyr Trp Ser Thr Gly Arg

Pro Pro Tyr

<210> 69

<211> 1558 <212> DNA

<213> Arabidopsis thaliana

<221> CDS <222> (191)..(1396) <223> G326

<400> 69

mbil9 Sequence Listing.ST25	60
caattaatga catcttcttc ttctcctttc actgcaaaac cgaaagcttg agactttgag attatgtcta tgtcatcttc ttcttcttcc atcgatcact tcatcacctt tcgtcatctt	120
gatettatte tecaetgtat aaaatcageg agattttaag ggattgtgaa ggtaccatet	180
tagacacaea ato got act tot act aca gag agt gtg gtg gcg tgt gaa	229
Met Gly Thr Ser Thr Thr Glu Ser Val Val Ala Cys Glu 1 5 10	
ttt tgc ggc gag aga acg gcg gtt ctg ttt tgt aga gcc gat acg gcg Phe Cys Gly Glu Arg Thr Ala Val Leu Phe Cys Arg Ala Asp Thr Ala 15 20 25	277
aag ctt tgt ttg cct tgt gac cag cac gtg cac tcg gcg aac ctt ctc Lys Leu Cys Leu Pro Cys Asp Gln His Val His Ser Ala Asn Leu Leu 30 35 40	325
tcg agg aag cat gtt cgt tct cag atc tgt gat aac tgt agc aaa gag Ser Arg Lys His Val Arg Ser Gln Ile Cys Asp Asn Cys Ser Lys Glu 50 55 60	373
ccg gtg tcc gta cgt tgc ttc aca gat aat ctc gta ttg tgt cag gag Pro Val Ser Val Arg Cys Phe Thr Asp Asn Leu Val Leu Cys Gln Glu 65 70 75	421
tgt gat tgg gat gtt cac gga agc tgt tcc tcc tcc gcg acg cat gaa Cys Asp Trp Asp Val His Gly Ser Cys Ser Ser Ser Ala Thr His Glu 80 85 90	469
cgc tcc gcc gtg gaa ggg ttt tca ggt tgt cct tcg gtt ttg gag ctt Arg Ser Ala Val Glu Gly Phe Ser Gly Cys Pro Ser Val Leu Glu Leu 95 100 105	517
gct gct gtg tgg gga atc gat tta aag ggt aag aag aaa gaa gat gac Ala Ala Val Trp Gly Ile Asp Leu Lys Gly Lys Lys Lys Glu Asp Asp 110 115 120 125	565
gaa gac gaa ttg act aag aat ttt ggg atg ggg ttg gat tcg tgg ggt Glu Asp Glu Leu Thr Lys Asn Phe Gly Met Gly Leu Asp Ser Trp Gly 130 135 140	613
tct gga tct aac atc gtt caa gaa ctg att gtt cct tat gat gtg tct Ser Gly Ser Asn Ile Val Gln Glu Leu Ile Val Pro Tyr Asp Val Ser 145 150 155	661
tgc aaa aag caa agc ttt agc ttt ggg agg tct aag cag gta gtg ttt Cys Lys Lys Gln Ser Phe Ser Phe Gly Arg Ser Lys Gln Val Val Phe 160 . 165 170	709
gaa cag ctt gag tta ctg aag aga ggc ttc gtt gaa ggc gaa gga gag Glu Gln Leu Glu Leu Leu Lys Arg Gly Phe Val Glu Gly Glu Gly Glu 175 180 185	757
att atg gtt ccg gag gga atc aat ggc gga gga agc att tct cag cca Ile Met Val Pro Glu Gly Ile Asn Gly Gly Gly Ser Ile Ser Gln Pro 190 200 205	805
tct ccg acg acg tcg ttt act tct ttg ctt atg tct caa agt ctt tgt Ser Pro Thr Thr Ser Phe Thr Ser Leu Leu Met Ser Gln Ser Leu Cys 210 215 220	853
ggt aat ggt atg caa tgg aat gct act aat cat agc act ggc cag aac Gly Asn Gly Met Gln Trp Asn Ala Thr Asn His Ser Thr Gly Gln Asn 225 230 235	901
act cag ata tgg gat ttt aac ttg gga cag tcg agg aac cct gat gaa Thr Gln Ile Trp Asp Phe Asn Leu Gly Gln Ser Arg Asn Pro Asp Glu 240 245 250	949
cct agt cca gtc gaa act aaa ggc tct act ttc aca ttc aac aac gtt Pro Ser Pro Val Glu Thr Lys Gly Ser Thr Phe Thr Phe Asn Asn Val 255 260 265	997
act cat ctc aag aac gat acc cga acc acc aat atg aat gct ttc aaa Page 91	1045

Thr	His	Leu	Lys	Asn	Asp	Thr	mbi Arq	19 S Thr	eque Thr	nce Asn	List Met	ing. Asn	ST25 Ala	Phe	Lys	
270					275					280					285	1093
gag Glu	agt Ser	tac Tyr	Gln	gag Glu 290	gat Asp	Ser	Val	His	Ser 295	Thr	Ser	Thr	Lys	Gly 300	Gln	1093
gaa Glu	aca Thr	tct Ser	aag Lys 305	agc Ser	aac Asn	aat Asn	att Ile	cct Pro 310	gct Ala	gcc Ala	att Ile	cac His	tcg Ser 315	cat His	aaa Lys	1141
agt Ser	tct Ser	aac Asn 320	gac Asp	tcc Ser	tgt Cys	ggc Gly	ttg Leu 325	cat His	tgc Cys	acg Thr	gaa Glu	cat His 330	att Ile	gct Ala	att Ile	1189
act Thr	agt Ser 335	aat Asn	aga Arg	gcc Ala	aca Thr	aga Arg 340	ttg Leu	gtg Val	gcg Ala	gta Val	acg Thr 345	aat Asn	gct Ala	gat Asp	cta Leu	1237
gag Glu 350	cag Gln	atg Met	gca Ala	cag Gln	aac Asn 355	aga Arg	gat Asp	aat Asn	gct Ala	atg Met 360	cag Gln	cgg Arg	tac Tyr	aag Lys	gaa Glu 365	1285
aag Lys	aag Lys	aaa Lys	acg Thr	cgg Arg 370	aga Arg	tat Tyr	gat Asp	aag Lys	acc Thr 375	ata Ile	aga Arg	tat Tyr	gaa Glu	acg Thr 380	agg Arg	1333
		aga Arg														1381
		gat Asp 400		tag	atgt	ctct	cc a	acgti	aggt	t ti	tacat	ttga	a gat	ccta	aagt	1436
tag	gaacı	ttt 1	tttg	gttt	t to	ctact	ttca	a act	acct	tgt	aaat	gtaa	at g	gatc	gatett	1496
cag	ctgca	ata a	atgt	gtggd	c ag	gattt	ttgt	: aat	tttt	acg	ttta	aacct	tc t	aaaa	aaaaa	1556
cag aa	ctgca	ata a	atgtg	gtggd	ec ag	gattt	ttgt	: aat	tttt	acg [°]	ttta	acct	tc t	aaaa	aaaaa	1556 1558
	0> 1 1> 4 2> 1	ata a 70 401 PRT Arabi					ttgt	: aat	tttt	:acg	ttta	acct	ctc t	caaaa	aaaaa	
<21 <21 <21	0 >	70 401 PRT					ttgt	: aat	ttt	:acg [·]	ttta	acct	te t	aaaa	aaaaa	
<21 <21 <21 <21 <40	0> 3 1> 4 2> 1 3> 2	70 401 PRT Arabi	idope	sis t	:hali	iana										
<pre></pre>	0> 3 1> 4 2> 1 3> 2 0> 3	70 401 PRT Arabi	idop: Ser	sis t Thr S	chali Thr	iana Glu	Ser	Val	Val 10	Ala	Сув	Glu	Phe	Cys 15	Gly	
<21 <21 <21 <21 <40 Met 1	0> 1 1> 4 2> 1 3> 2 0> 3 Gly	70 401 PRT Arabi 70	Ser Ala 20	Thr S Val	thali Thr Leu	iana Glu Phe	Ser Cys	Val Arg 25	Val 10 Ala	Ala Asp	Cys Thr	Glu Ala	Phe Lys 30	Cys 15 Leu	Gly Cys	
aa <21 <21 <21 <21 <40 Met 1	00> 11> 4 12> 11 33> 11 Gly Arg	70 401 PRT Arabi 70 Thr	Ser Ala 20 Asp	Thr 5 Val	Thr Leu His	Glu Phe Val	Ser Cys His 40	Val Arg 25 Ser	Val 10 Ala	Ala Asp Asn	Cys Thr Leu	Glu Ala Leu 45	Phe Lys 30 Ser	Cys 15 Leu Arg	Gly Cys Lys	
aa <21 <21 <21 <400 Met 1	00> 11> 4 12> 11> 4 23> 17 33> 17 Gly Pro Val 50	70 401 PRT Arabi 70 Thr Cys 35	Ser Ala 20 Asp	Thr 5 Val Gln	Thr Leu His	Glu Phe Val Cys 55	Ser Cys His 40	Val Arg 25 Ser Asn	Val 10 Ala Ala Cys	Ala Asp Asn Ser	Cys Thr Leu Lys 60	Glu Ala Leu 45 Glu	Phe Lys 30 Ser	Cys 15 Leu Arg	Gly Cys Lys Ser	
aa <21 <21 <21 <21 <40 Met 1 Glu His Val 65	0> 11> 4 11> 4 12> 11 13> 10 14 150 150 150 150 150 150 150 150 150 150	70 401 PRT Arabi 70 Thr Thr Cys 35	Ser Ala 20 Asp	Thr 5 Val Gln Gln Thr	Thr Leu His Ile Asp	Glu Phe Val Cys 55 Asn	Ser Cys His 40 Asp	Val Arg 25 Ser Asn	Val 10 Ala Ala Cys	Ala Asp Asn Ser Cys 75	Cys Thr Leu Lys 60 Gln	Glu Ala Leu 45 Glu Glu	Phe Lys 30 Ser Pro	Cys 15 Leu Arg Val	Gly Cys Lys Ser Trp 80	

# mbil9 Sequence Listing.ST25

Trp	Gly	Ile 115	Asp	Leu	Lув	Gly	Lys 120	Lys	Lys	Glu	Asp	Авр 125	Glu	Asp	Glu
Leu	Thr 130	Lys	Asn	Phe	Gly	Met 135	Gly	Leu	Asp	Ser	Trp 140	Gly	Ser	Gly	Ser
Asn 145	Ile	Val	Gln	Glu	Leu 150	Ile	Val	Pro	Tyr	Asp 155	Val	Ser	Сув	Lys	Lys 160
Gln	Ser	Phe	Ser	Phe 165	Gly	Arg	Ser	Lys	Gln 170	Val	Val	Phe	Glu	Gln 175	Leu
Glu	Leu	Leu	Lys 180	Arg	Gly	Phe	Val	Glu 185	Gly	Glu	Gly	Glu	Ile 190	Met	Val
Pro	Glu	Gly 195	Ile	Asn	Gly	Gly	Gly 200	Ser	Ile	Ser	Gln	Pro 205	Ser	Pro	Thr
Thr	Ser 210	Phe	Thr	Ser	Leu	Leu 215	Met	Ser	Gln	Ser	Leu 220	Сув	Gly	Asn	Gly
Met 225	Gln	Trp	Asn	Ala	Thr 230	Asn	His	Ser	Thr ·	Gly 235	Gln	Asn	Thr	Gln	Ile 240
Trp	Авр	Phe	Asn	Leu 245	Gly	Gln	Ser	Arg	Asn 250	Pro	Авр	Glu	Pro	Ser 255	Pro
Val	Glu	Thr	Lys 260	Gly	Ser	Thr	Phe	Thr 265	Phe	Asn	Asn	Val	Thr 270	His	Leu
Lys	Asn	Asp 275	Thr	Arg	Thr	Thr	Asn 280	Met	Asn	Ala	Phe	Lys 285	Glu	Ser	Туг
Gln	Glu 290	Asp	Ser	Val	His	Ser 295	Thr	Ser	Thr	Lys	Gly 300	Gln	Glu	Thr	Ser
Lys 305	Ser	Asn	Asn	Ile	Pro 310	Ala	Ala	Ile	His	Ser 315	His	Lys	Ser	Ser	Asn 320
Asp	Ser	Сув	Gly	Leu 325	His	Сув	Thr	Glu	His 330	Ile	Ala	Ile	Thr	Ser 335	Asn
Arg	Ala	Thr	Arg 340	Leu	Val	Ala	Val	Thr 345	Asn	Ala	Asp	Leu	Glu 350	Gln	Met
Ala	Gln	Asn 355	Arg	Asp	Asn	Ala	Met 360	Gln	Arg	туг	Lys	Glu 365	Lys	Lys	Lys
Thr	Arg 370	Arg	Tyr	Asp	Lys	Thr 375	Ile	Arg	Tyr	Glu	Thr 380	Arg	Lys	Ala	Arg
Ala 385	Glu	Thr	Arg	Leu	Arg 390	Val	Lys	Gly	Arg	Phe 395	Val	Lys	Ala	Thr	Asp 400

Pro

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/31414

A. CLASSIFICATION OF SUBJECT MATTER  IPC(7) : A01H 1/00, 5/00; C12N 5/04, 15/00, 15/82; C12P 21/02				
US CL : 435/69.1, 320.1, 410, 419, 468, 800/278, 284, 287, 290				
According to International Patent Classification (IPC) or to both national classification and IPC  B. FIELDS SEARCHED				
		hy classification symbols)		
Minimum documentation searched (classification system followed by classification symbols) U.S.: 435/69.1, 320.1, 410, 419, 468; 800/278, 284, 287, 290				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched				
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) East, USPAT; STN, Agricola, Biosis, CaPlus, Embase; Sequence Search of SEQ ID NOs. 1 & 2				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category *	Citation of document, with indication, where ap	propriate, of the relevant passages	Relevant to claim No.	
X	AOYAMA, T. et al. Ectopic expression of the Arab 1 alters leaf cell fate in tobacco, The Plant Cell, No	vember 1995. Vol. 7. pages 1773-	1-10, 13, 25	
Y	1785, entire document.		11, 12, 26	
x	BELLIS, L.D. et al. Distinct cis-acting sequences a sugar responses of the cucumber isocitrate lyase gen 378, entire document.	re required for the germination and te. Gene 1997, Vol. 197, pages 375-	1-10, 13, 25	
X	KIM, S. et al. Sugar response element enhances wor inhibitor II promoter in transgenic tobacco. Plant M 983, entire document.	und response of potato proteinase ol. Biol. 1991, Vol. 17, pages 973-	1-10, 13, 25	
Y	Database Genbank on NCBI, US National Library on No. U78721, LIN, X. et al. 'Sequence and analysis Arabidopsis thaliana' 5 April 2000, especially bases	of chromosome 2 of the plant	1-10, 13, 25	
Further documents are listed in the continuation of Box C. See patent family annex.				
"A" document defining the general state of the art which is not considered to be of particular relevance		"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention		
"E" earlier application or patent published on or after the international filing date		considered novel or cannot be considered	red to involve an inventive step	
"L" document which may throw doubts on priority claim(s) or which is cited to		"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is		
"O" document referring to an oral disclosure, use, exhibition or other means		combined with one or more other suc being obvious to a person skilled in th		
	t published prior to the international filing date but later than the late claimed	"&" document member of the same patent	family	
Date of the actual completion of the international search  Date of mailing of the international search report  3. Rehmany 2001 (23.02.2001)  0.4 APR 2001			arch report	
	2001 (23.02.2001) nailing address of the ISA/US	Authorized officer	mx)	
Commissioner of Patents and Trademarks			AT DEA ANA	
	PCT shington, D.C. 20231	PINILOS		
Facsimile No. (703)305-3230  Telephone No. 703-308-TEGENOLOGY CENTER 1600				

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/31414

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)			
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:			
	Claim Nos: because they relate to subject matter not required to be searched by this Authority, namely:		
	Claim Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:		
.4(a).	Claim Nos.: 14 & 24 because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule		
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)			
nternati	onal Searching Authority found multiple inventions in this international application, as follows: ontinuation Sheet		
	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:		
ork on	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-13, 25 and 26; SEQ ID NOs 1 & 2  Protest		
	A(a). Obsternation		

### INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/31414

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups I-XXXV, claim(s) 1-13 and 25-26, drawn to a transgenic plant having modified seed characteristics, polynucleotides and vectors for producing said transgenic plant and a method of making said transgenic plant. Applicant must elect one pair of sequences (one nucleic acid and the corresponding amino acid translation) to be examined, i.e. SEQ ID NO: 1 and 2 in Group I, SEQ ID NO: 3 and 4 in Group II, SEQ ID NO: 5 and 6 in Group III, etc.

Group XXXVI, claim(s) 15-17, drawn to a method of identifying a factor that is modulated.

Group XXXVII, claims(s) 18, drawn to a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide.

Group XXXVIII, claims(s) 19 and 20, drawn to an integrated computer system.

Group XXXIV, claim(s) 21-23, drawn to a method for identifying a polynucleotide sequence comprising selecting a nucleic acid sequence from a database that meets a selected sequence criteria.

The inventions listed as Groups I-XXXIV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions listed as Groups I-XXXIX do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I-XXXV are drawn to a transgenic plant and a method of producing said plant with a nucleic acid sequence. The methods of Groups I-XXXV differ from each other in that they are directed to a plant transformation method and transgenic plant with a structurally and functionally distinct nucleic acid sequence which encodes a structurally and functionally distinct amino acid sequence. In addition, Groups XXXVI, XXXVII and XXXIX are different methods from any of Groups 1-XXXV in that they have different method steps and different end products, and Group XXXVIII requires a computer system. Thus, there is no single special technical feature, which links the inventions of Groups 1-XXXIX under PCT Rule 13.2.